

JQ

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
12 December 2002 (12.12.2002)(10) International Publication Number  
WO 02/099078 A2

PCT

(51) International Patent Classification<sup>7</sup>: C12N

(21) International Application Number: PCT/US02/18065

(22) International Filing Date: 5 June 2002 (05.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/296,525 5 June 2001 (05.06.2001) US  
60/363,901 11 March 2002 (11.03.2002) US

(71) Applicant (for all designated States except US): IRM LLC [US/US]; P.O. Box HM 2899, HM LX Hamilton (BM).

(72) Inventors; and

(75) Inventors/Applicants (for US only): WINSSINGER, Nicolas [BE/US]; 643 Gravilla Street, La Jolla, CA 92037 (US). HARRIS, Jennifer, L. [US/US]; 4572 Tivoli

Street, San Diego, CA 92107 (US). BACKES, Bradley, J. [US/US]; 1901A Oxford Avenue, Cardiff by the Sea, CA 92007 (US). SCHULTZ, Peter, G. [US/US]; 1650 La Jolla Rancho, La Jolla, CA 92037 (US).

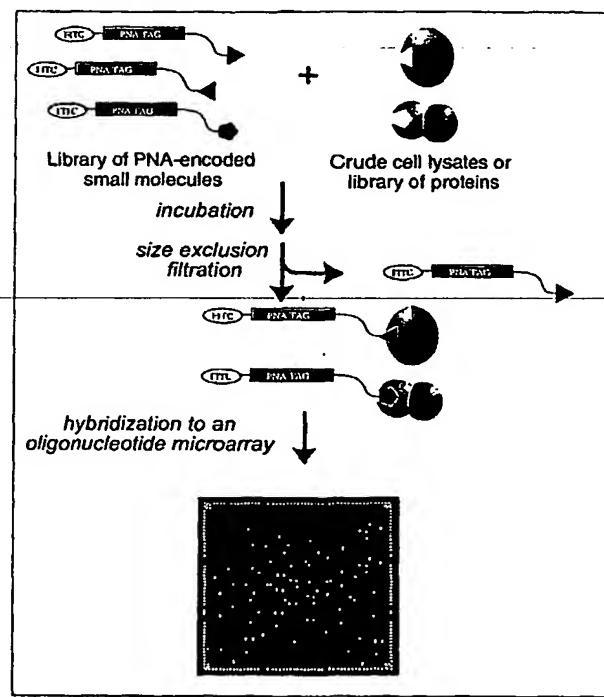
(74) Agents: GARRETT-WACKOWSKI, Eugenia et al.; TOWNSEND AND TOWNSEND AND CREW LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

## (54) Title: FUNCTIONAL PROTEOMIC PROFILING



(57) Abstract: Oligonucleotides such as PNA are used to code for the identity of individual members within a library such that a library present as a mixture can be converted into a spatially addressable format through hybridization to an oligonucleotide array. Such a strategy can greatly facilitate arraying small molecules, antibodies, proteins and oligosaccharides on a chip and allows for binding or other assays to be performed in solution prior to hybridization. This invention is particularly useful for chemical libraries as it allows for combinatorial synthesis employing split and mix technology with positional encoding. The decoding is achieved by hybridization to an oligonucleotide. With high density oligonucleotide arrays, every library member can be analyzed in a highly miniaturized format. As such, this technology readily lends itself to highly miniaturized screening of single or multiple targets simultaneously and profiling.

WO 02/099078 A2

BEST AVAILABLE COPY



Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

**Declaration under Rule 4.17:**  
— *of inventorship (Rule 4.17(iv)) for US only*

**Published:**

— *without international search report and to be republished  
upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## FUNCTIONAL PROTEOMIC PROFILING

### CROSS-REFERENCE TO RELATED APPLICATIONS

[01] The present provisional patent application claims priority to U.S. Provisional Patent Application Nos. 60/296,525, filed on June 5, 2001, and 60/363,901, filed March 11, 2002, the teachings of both of which are incorporated herein by reference for all 5 purposes.

### FIELD OF THE INVENTION

[02] This invention pertains to the field of proteomic profiling and encoding the synthesis of compounds to facilitate the identification of particular compounds in a library.

10

### BACKGROUND OF THE INVENTION

[03] Numerous technologies have been developed to investigate cellular events on a genome-wide scale. Oligonucleotide arrays provide information on changes in mRNA expression levels in response to a variety of physiological stimuli (see, Lockhart, *et al.*, *Nature Biotechnol.*, 14:1675 (1996); and DeRisi, *et al.*, *Science*, 278:680 (1997); and 15 Lockhart and Winzeler, *Nature*, 405:827 (2000)). Two-dimensional gel electrophoresis, or other chromatographic separation methods, in conjunction with mass spectroscopy offer a more direct analysis of proteome function (see, Anderson and Anderson, *Electrophoresis*, 19:1853 (1998); Figeys, *et al.*, *Nat. Biotechnol.*, 11:1544 (1996); and for reviews, see: 20 Cortahs, *et al.*, *Electrophoresis*, 21:1104 (2000); and Gygi, *et al.*, *Proc. Natl. Acad. Sci. USA*, 97:9390 (2000)). Technologies have also been developed for genome-wide analysis of protein structure (see, Abola, *et al.*, *Nat. Struct. Biol.*, 7:973 (2000)) Stevens, *Curr. Opin. Struct. Biol.*, 10:558 (2000)). In a more targeted analysis of protein function, maps of protein-protein and protein-DNA interactions have been reported as well as preliminary work 25 towards a protein chip (see, Uetz, *et al.*, *Nature*, 403:623 (2000); Iyer, *et al.*, *Nature*, 409:533 (2001); Zhu, *et al.*, *Nature Genet.*, 26:283 (2000); Arenkov, *et al.*, *Anal. Biochem.*, 278:123 (2000); and MacBeath and Schreiber, *Science*, 289:2760 (2000)). Methods to monitor the catalytic activity of proteins on a genome-wide scale also provide critical insights into 30 cellular activity (see, Goulet, *J. Gen. Microbiol.*, 87:97 (1975); Kam, *et al.*, *Bioconjug. Chem.*, 4:560 (1993); Abuelayan, *et al.*, *Bioconjug. Chem.*, 5:400 (1994); Liu, *et al.*, *Proc. Natl. Acad. Sci. USA*, 96:14694 (1999); Greenbaum, *et al.*, *Chem. Biol.*, 8:569 (2000);

Nicholson, *et al.*, *Nature*, 376:37 (1995); Backes, *et al.*, *Nat. Biotechnol.*, 18:187 (2000); and Nazif, *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, 98:2967-72 (2001)).

[04] Small molecules have long been used to analyze and control the catalytic activity of enzymes as well as modulate biological networks by acting as agonists or antagonists of receptors (see, Gray, *et al.*, *Science*, 281:533 (1998); and Hung, *et al.*, *Chem Biol*, 3:623 (1996)). As such, microarrays of small molecule inhibitors or substrates provide a tool for profiling cellular activity. If one hopes to discriminate between the >30,000 potential gene products in humans, it is clear that microarrays containing large collections of compounds will be a necessity. High density microarrays of peptides and unnatural oligomers have been reported (40,000 compounds/cm<sup>2</sup>), however the photolithographic techniques used limit the range of accessible molecular diversity (see, Fodor, *et al.*, *Science*, 251:767 (1991); Cho, *et al.*, *Science*, 261:1303 (1993)). More recently, several small molecules have been printed on a glass slide in an effort to merge robotic printing and split-pool libraries (see, MacBeath and Schreiber, *J. Am. Chem. Soc.*, 121:7967 (1999); and Hergenrother, *et al.*, *J. Am. Chem. Soc.*, 122:7849 (2000)). Split-pool library synthesis is far more efficient for the generation of molecular diversity than parallel synthesis, as the number of final products in a split-pool library is exponentially related to the number of diversity introducing reactions, whereas it is linearly proportional in parallel synthesis (see, Furka, *et al.*, *Highlights of Modern Biochemistry, Proceedings of the 14th International Congress of Biochemistry, Prague, Czechoslovakia, 1988*; VSP: Utrecht, The Netherlands,; 13:47-47 (1988); Furka, *et al.*, *Int. J. Pept. Protein Res.*, 37:487 (1991); Lam, *et al.*, *Nature*, 354:82 (1991); and Houghton, *et al.*, *Nature*, 354:84 (1991)). However, in split-pool synthesis the identity of each library member is unknown and must be individually decoded for each active library member (see, Brenner and Lerner, *Proc. Natl. Acad. Sci. U.S.A.*, 89:5381 (1992); and Needels, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 90:10700 (1993)). If one wishes to screen such libraries against >30,000 gene products, the decoding of library members becomes problematic.

[05] Thus, a need exists for technologies for screening and decoding of large numbers of chemically diverse library members. The present invention fulfils this and other needs.

#### SUMMARY OF THE INVENTION

[06] The present invention provides a novel strategy for encoding the identity of synthesized molecules. The methods utilize a stable and easily synthesized PNA (peptido nucleic acid, Figure 1) tag which is tethered to the small molecule to code for its

structure. The PNA tag serves two purposes: first, to encode the synthetic history of the small molecule, and second, to positionally encode the identity of the small molecule by its location upon hybridization to an oligonucleotide microarray. The methodology provided by the present invention thus avoids the two biggest limitations of the previously used split-and-pool combinatorial synthesis methods.

[07] In one embodiment, the present invention provides a method for preparing a library of diverse compounds, each of the compounds being produced by the step-by-step assembly of building blocks, the method comprising the steps of: (a) apportioning solid supports among a plurality of reaction vessels; and (b) in each reaction vessel of the plurality of reaction vessels, exposing the solid supports to a first building block of a compound and to a first monomer of a peptido nucleic acid (PNA) identifier tag under conditions suitable for immobilization of the first building block and the first monomer, wherein the first building block present in one reaction vessel is different from the first building block present in at least one of the other reaction vessels, wherein the first building block of the compound is capable of being covalently coupled to a second building block and wherein the first monomer of the PNA identifier tag is capable of being covalently coupled to a second monomer. In one embodiment, the method further comprises: (c) pooling the solid supports. In another embodiment, the method further comprises: (c) cleaving the first compound from the solid support. In some embodiments, the first building block of the first compound is an amino acid. Suitable amino acids include, but are not limited to, the following: L-amino acids, D-amino acids,  $\alpha$ -amino acids,  $\beta$ -amino acids and  $\omega$ -amino acids.

[08] In some embodiments, the methods further comprise: (d) reapportioning the pooled solid supports among a plurality of reaction vessels; and, (e) in each reaction vessel of the plurality of reaction vessels, exposing the solid supports to at least a second building block of the compound and to at least a second monomer of the PNA identifier tag under conditions suitable for attachment of the second building block to the first building block of the compound and the second monomer to the first monomer of the PNA identifier tag, wherein the second building block present in one reaction vessel is different from the second building block present in at least one of the other reaction vessels.

[09] In additional embodiments, the solid supports that are apportioned in (a) each further comprise at least a third building block and at least a third monomer of a PNA identifier tag, wherein the third monomer of the PNA identifier tag identifies the third building block, and wherein the first building block attaches to the third building block and the first monomer attaches to the third monomer.

[10] The PNA identifier tag(s) can be of a variety of lengths. In a preferred embodiment, the PNA identifier tag, *e.g.*, the first PNA identifier tag, is from about 3 to about 50 nucleotides in length. In another embodiment, the PNA identifier tag is from about 6 to about 20 nucleotides in length. In yet another embodiment, the PNA identifier tag is about 12 nucleotides in length. In another preferred embodiment, the PNA identifier tag, *e.g.*, the first identifier tag, further comprises a label. Suitable labels include, but are not limited to, fluorophores, radioactive labels, *etc.*

[11] In one embodiment, the first building block is immobilized on the solid support. In another embodiment, the first monomer is immobilized on the solid support. In another embodiment, the first monomer is immobilized on the first building block and not on the solid support. In another embodiment, the first building block is immobilized on the solid support and the first monomer is immobilized on the first building block. In certain preferred embodiments, the first monomer is immobilized on the first building block through a linker.

[12] Numerous solid supports can be used in the methods of the present invention. In some embodiments, the solid support is a bead or particle. In other embodiments, the solid support is a nonporous bead. In certain preferred embodiments, the solid support is a bead having a diameter ranging from about 1 nm to about 1 mm.

[13] In some embodiments, prior to exposing the first building block to the solid support, the first building block is activated to facilitate immobilization of the first building block onto the solid support. In other embodiments, prior to exposing the first monomer to the solid support, the first monomer is activated to facilitate immobilization of the first monomer onto the solid support. In other embodiments, prior to exposing the first monomer to the solid support, the first monomer is activated to facilitate immobilization of the first monomer onto the first building block. In other embodiments, the solid support is exposed to the first monomer after the solid support is exposed to the first building block.

[14] In preferred embodiments, steps (a) through (c) are carried out so as to construct a library of at least 10 different compounds. In other embodiments, steps (a) through (c) are carried out so as to construct a library of at least 100 different compounds. In other embodiments, steps (a) through (c) are carried out so as to construct a library of at least  $10^3$  different compounds. In other embodiments, steps (a) through (c) are carried out so as to construct a library of at least  $10^4$  different compounds. In other embodiments, steps (a) through (c) are carried out so as to construct a library of at least  $10^5$  different compounds. In other embodiments, steps (a) through (c) are carried out so as to construct a library of at least  $10^6$  different compounds.

[15] In another embodiment, the present invention provides a method for identifying a compound that binds a target, the method comprising: (a) contacting the target with a library of compounds, wherein each of the compounds comprises a peptido nucleic acid (PNA) identifier tag; (b) separating the compounds (and/or a detectable label attached to the compound) that bind the target from those compounds (and/or labels) that do not bind the target to obtain target-compound complexes; (c) hybridizing the target-compound complexes to an array of oligonucleotides; and (d) detecting the target-compound complexes that hybridize to the array of oligonucleotides, thereby identifying the compounds that bind the target.

[16] In one embodiment, the target is a protein. In some embodiments, the target can be in a cell extract, a tissue, a biological sample, a sample from an industrial process and the like. In some embodiments, the target comprises a label. Suitable labels include, but are not limited to, fluorophores, radioactive labels, *etc.* In another embodiment, the target is a library of targets. In another embodiment, each of the compounds further comprises a label. In some embodiments, the label is attached to the PNA identifier tag. Again, suitable labels include, but are not limited to, fluorophores, radioactive labels, *etc.*

[17] Numerous methods can be used to carry out step (b) of the above method. Typically, any method that is capable of separating the compounds that bind the target from those compounds that do not can be used. In a preferred embodiment, step (b) is carried out using, for example, size-exclusion chromatography or affinity chromatography. It will be readily apparent to those of skill in the art that other separation techniques can be used to carry out step (b).

[18] In step (c) of the above method, the target-compound complexes are hybridized to an array of oligonucleotides. Numerous different oligonucleotide arrays can be employed. An example of one such oligonucleotide array is the GenFlex™ tag array, which is commercially available from Affymetrix (Santa Clara, California). In one embodiment, each of the oligonucleotides in the array is about 10 to about 50 nucleotides in length. In another embodiment, each of the oligonucleotides in the array is about 20 to about 30 nucleotides in length. In some embodiments, the PNA identifier tag hybridizes to the terminal portion of the oligonucleotides in the oligonucleotide array.

[19] In another embodiment, the present invention provides a method for identifying a compound that binds a target, the method comprising: (a) providing a library of compounds, wherein each of the compounds comprises a peptido nucleic acid (PNA) identifier tag; (b) hybridizing the library of compounds to an array of oligonucleotides; (c)

contacting the array of bound compounds with a target; and (d) detecting compounds that bind the target.

[20] In one embodiment of the above method, the target is a protein, such as an enzyme. In some embodiments, the target comprises a label. Suitable labels include, but are not limited to, fluorophores, radioactive labels, *etc.* In some embodiments, the target is a library of targets. In this embodiment, each of the different targets preferably comprises a different label (*e.g.*, a different fluorophore).

[21] In some embodiments, for example, when one seeks to detect an enzyme that cleaves a particular compound, the PNA-tagged compounds comprise one or more labels. To determine whether a target can cleave the compound, the library of targets is contacted with the compounds which, in turn, are hybridized to an oligonucleotide array. The presence or absence of the label at particular positions is then indicative of whether the target has cleaved (and therefore released the label from) the compound that is immobilized at that location of the array.

[22] Other features, objects and advantages of the invention and its preferred embodiments will become apparent from the detailed description, examples, claims and figures that follow.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[23] Figure 1 illustrates the chemical structures of DNA and PNA.

[24] Figure 2 illustrates the chemical structure of PNA, protected with orthogonal protecting groups such as Fmoc, Bmoc, or Alloc (P and P<sup>2</sup>).

[25] Figure 3 illustrates a schematic of split pool synthesis of PNA encoded combinatorial libraries. R = element of diversity present in library, B = base of the peptidomimetic acid, x = number of base encoding a single element of diversity, n = number of chemical diversification steps, P = protecting group.

[26] Figure 4 illustrates an example of split-and-pool combinatorial synthesis. A1 through A3 represent building blocks used to introduce functional diversity into a combinatorial library. The introduction of diversity is accompanied by an encoding step, wherein each library member is derivatized with a tag that will be used to identify the each library member.

[27] Figures 5A and B illustrate two different formats for screening using PNA-encoded libraries. In Figure 5A, the PNA tagged-compound is not labeled. In Figure 5B, several possible arrangements are shown in which the PNA-tagged compound is labeled

with one or more label moieties. Although the label moieties are shown as being fluorophores, other labels are also suitable.

[28] Figure 6 illustrates an example of proteomic profiling using a library of PNA-encoded small molecules. Individual small molecules are tethered to a unique PNA sequence which encodes both their synthetic history and their location upon hybridization to an oligonucleotide microarray. The PNA are capped with fluorescein for fluorescence detection. The library of PNA-encoded small molecule is incubated with a protein mixture of interest, passed through a size exclusion filter to separate the small molecules-PNA adducts bound to a macromolecule from the unbound ones and the high molecular weight fraction is hybridized to the oligonucleotide microarray.

[29] Figure 7 illustrates a scheme for the split-and-pool synthesis of a PNA encoded combinatorial library (Scheme 1).

[30] Figure 8 illustrates the synthesis of protected C, T, A, and G PNA monomers. The process involves 17 synthetic steps.

[31] Figure 9 illustrates split-and-pool synthesis of a PNA-encoded library for kinase profiling. All amino acid side chain residues and PNA monomers are protected with acid labile groups. Using presynthesized codons, the library requires 22 synthetic operations.

[32] Figure 10 illustrates a scheme for the synthesis of a designed cathepsin C inhibitor with and without a PNA tag. Fmoc = 9-fluorenylmethoxycarbonyl; Alloc = allyloxycarbonyl.

[33] Figure 11 illustrates the chemical structures of designed PNA-tagged cysteine protease inhibitors 3-8. FITC = fluorescein thiocarbamate.

[34] Figure 12A illustrates the hybridization of probes 3-8 from Figure 11 (45 pmol of each probe was hybridized). Figure 12B is a control in which probes 3-8 (1.4  $\mu$ M) were incubated for 2 hour at pH 5.5, subjected to size exclusion chromatography, and hybridization to the array. Figure 12C illustrates the results of an incubation of probes 3-8 with cathepsin C (100  $\mu$ M, 20  $\mu$ l) for 2 hours at pH 5.5, followed by size exclusion chromatography and hybridization to the array. Figure 12D illustrates the results of an incubation of probes 3-8 (1.4  $\mu$ M) with cathepsin L (10  $\mu$ M) for 2 hours at pH 5.5, followed by size exclusion chromatography and hybridization to the array.

[35] Figure 13 illustrates a PNA encoded-library for protease profiling using a FRET reporting system. The library is synthesized using a similar protocol to the

kinase library as shown in Figure 9. The fluorophores are protected with acid labile groups. Amplification was performed using anti-fluorescein antibody, biotinylated secondary antibody, and phycoerythrin-labeled streptavidin.

5 [36] Figure 14 illustrates the chemical structure of selected protease inhibitors 1-7.

[37] Figure 15 illustrates the quantification of protease activity. A. Probes 1-7 (20  $\mu$ L at 1.0  $\mu$ M) were incubated with various concentration of caspase-3 (10-500 nM), passed through a size exclusion filter and hybridized to an GenFlex™ (www.affymetrix.com) microarray (False color scale, the probes at the top of the image are control probes. The 10 intensity has been standardized in the five images for qualitative viewing purposes). B. Correlation of protease activity and observed probe intensity. Plot of the fluorescence intensity (X-axis) vs. caspase-3 concentration (Y-axis) with a standard error of 10%.

15 [38] Figure 16 illustrates crude cell lysate profiling. A. Direct hybridization of compound 1-7; B. incubation of compound 1-7 with granzyme B, size exclusion, hybridization; C. Incubation of compound 1-7 with purified caspase-3, size exclusion, hybridization; D. Incubation of compound 1-7 with Jurkat crude cell lysate, size exclusion, hybridization; E. Incubation of compound 1-7 with crude cell lysate from Jurkat cells pretreated with granzyme B, size exclusion, hybridization.

20 [39] Figure 17 illustrates MS/MS spectra of (A) triply charged and (B) doubly charged SGTDVDAANLRETFR peptides derived from human caspase-3. Prominent y92+ and y112+ ions in the MS/MS spectrum of the doubly charged precursor (B) are consistent with facile cleavage of the C-terminal to aspartic acid residues reported in the literature<sup>24</sup>. Loss of the elements of water is denoted by \*, while ammonia loss is indicated by #.

25 [40] Figure 18 illustrates inhibition of the apoptosis phenotype. A.

Inhibition of downstream caspase-3 mediated autoprocessing and cleavage of DFF-45 upon incubation of granzyme B activated Jurkat lysates with inhibitor 6c. The blots were probed with Anti-Caspase 3 and Anti-DFF45 and then visualized. B. Inhibition of fas-mediated apoptosis by caspase inhibitor (Z-D(Ome)-E(Ome)-V-D(Ome)-FMK). Cells were stained 30 with both Anexin-V-EGFP and Propidium Iodide.

[41] Figure 19 sets forth examples of "war-heads" that are suitable for use in protease profiling.

**DETAILED DESCRIPTION OF THE INVENTION  
AND PREFERRED EMBODIMENTS**

**A. General Overview**

[42] The present invention provides a novel strategy for encoding the identity of synthesized molecules. Combinatorial libraries, for example, are often synthesized using a "split-and-pool" strategy. This process has several shortcomings, including the difficulty in determining the structure of those library members that have a desired activity, *e.g.*, binding to a particular ligand, inhibition of enzymatic activity, and the like. The present invention solves these problems by combining the spatial addressability that is obtainable using arrays of oligonucleotides with split-and-pool synthesis. This technology is particularly well-suited for the screening of multiple enzymes against a library of molecules and readily extends to screening of libraries of proteins, such as the proteome of crude cell extracts, against libraries of compounds, such as small organic molecules.

[43] The methods of the present invention involve the preparation of microarrays of molecules using positionally encoded libraries. One aspect of the novelty of the methods of the present invention is the hybridization of the library to a spatially addressable oligonucleotide array, *e.g.*, a DNA chip, thereby reformatting the split-and-pool library into a spatially addressable one. The methodology provided by the present invention thus avoids the two biggest limitations of split-and-pool combinatorial synthesis. First, the screening can be performed in solution (*i.e.*, the hybridization to the DNA chip can be carried out after incubation of the library with the enzyme(s) or other target compound(s) of interest). Second, the decoding step is virtually instantaneous (scanning a 400,000 features DNA chip requires less than 5 minutes) and is independent of the number of hits. Additionally, a time consuming inconvenience associated with split-and-pool screening is that more than one bead per compound must be used in order to ensure that all library members are present. Thus, decoding of active beads often gives redundant results.

[44] Although others have proposed the use of oligonucleotide tags, these previously described methods used a decoding by PCR amplification of the polymer-bound tag (an approach similar to the more popular haloaromatic tags developed by Still *et al.*). Thus, these methods suffer from the same limitations as other encoding methods. Furthermore, oligonucleotide tags are very limiting in terms of the chemistry that can be used to construct the small molecule libraries, whereas PNA tags are not.

[45] In preferred embodiments, the methods utilize a stable and easily synthesized PNA (peptido nucleic acid, Figure 1) tag which is tethered to the small molecule

to code for its structure. The PNA tag serves two purposes: first, to encode the synthetic history of the small molecule, and second, to positionally encode the identity of the small molecule by its location upon hybridization to an oligonucleotide microarray.

[46] As mentioned above, there are numerous advantages associated with the use of PNA tags. For instance, PNAs (Figure 1) are particularly suitable as the encoding oligonucleotides based on their desirable hybridization properties, the flexibility of their synthesis and their chemical robustness (PNAs are compatible with 95% TFA used in numerous cleavages). For library synthesis, the oligomerization of PNAs relies on an amide bond formation, one of the mildest, most reliable and versatile reactions in organic chemistry.

5 It can be performed under neutral, acidic or basic conditions and there is a wide variety of known protecting groups to mask the nitrogen of each monomer thus insuring that the chemistry of the oligonucleotide is compatible with the widest array of chemistry for library synthesis. The wide array of possible protecting groups for the nitrogen of the PNA's N-terminus can accommodate a wide range of diversity introducing reactions. For example, one

10 15 can mask the nitrogen of the PNA as an azide or an allyl carbamate (Alloc) based on their mildness of unmasking and their stability. Finally, in terms of hybridization properties, the lack of negative charges on the PNA backbone increases its affinity for DNA and reduces the influence of salt concentration on hybridization strength. Unlike DNA-DNA interactions, PNA-DNA interactions are fairly insensitive to sodium ion concentration and thus offer more

20 flexibility in the choice of buffer systems for screening purposes.

**B. Definitions**

[47] All technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The present definitions and abbreviations are generally offered to supplement the art-recognized meanings. Generally, the nomenclature used herein and the laboratory procedures organic chemistry, peptide synthesis and enzyme chemistry described below are those well known and commonly employed in the art. Generally, enzymatic reactions and purification steps are performed according to the manufacturer's specifications. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses.

30 [48] The term "substrate" or "solid support" refers to a material having a rigid or semi-rigid surface which contains or can be derivatized to contain reactive functionality that covalently links a target compound or a PNA identifier tag to the surface thereof. Such materials are well known in the art and include, by way of example, silicon

dioxide supports containing reactive Si-OH groups, polyacrylamide supports, polystyrene supports, polyethyleneglycol supports, and the like. Such supports will preferably take the form of small beads, pellets, disks, or other conventional forms, although other forms may be used. In some embodiments, at least one surface of the substrate will be substantially flat. In 5 preferred embodiments, the substrate or solid support is roughly spherical.

[49] The term "reactions" refers to any reaction that adds a monomer to the solid support, that modifies the chemical entity formed after monomer addition to the solid support and/or that removes a group from the solid support. The reactions can employ monomers (building blocks) that become incorporated onto the solid support or can merely 10 employ a reagent, such as heat, base, acid, an oxidizing agent, a reducing agent, an enzyme, *etc.* that does not become incorporated into the structures found on the support.

Modifications of the chemical entity formed after monomer addition to the solid support include, for example, cyclization, isomerization, *etc.* Removal of a group from the solid support includes hydrolysis to remove an ester, removal of protecting groups, *etc.*

[50] The term "target compound" refers to the compound or a group of compounds to be synthesized on the solid support and subsequently screened for biological activity or other properties, either on the solid support or after it has been removed from the solid support. The term "target compound" is used interchangeably herein with the terms "oligomer," "polymer," and "small molecule."

[51] The term "monomer(s)" as used relative to target compound synthesis or PNA identifier tag synthesis refers to discreet building blocks employed to prepare the target compound or the PNA identifier tag. Thus, in the case of thiazolidone compound synthesis on a solid support by reaction of an amine, an aldehyde and a thioacetic acid compound, each of the amine, aldehyde and thioacetic acid is a monomer in the synthesis of 25 the thiazolidone. In the case of peptide synthesis, the monomer is typically an amino acid, but can comprise a di- or higher amino acid fragment of the target peptide that is incorporated as a single entity. In the case of PNA identifier tag synthesis, the monomer is a nucleotide or a string of nucleotides. The term "monomer(s)" is used interchangeably with the term "building block(s)," and both terms are used in connection with the synthesis of the target 30 compound as well as the synthesis of the peptido nucleic acid (PNA) identifier tag.

[52] The term "peptido nucleic acid identifier tag" or "PNA identifier tag" or "PNA tag" refer to a PNA sequence that serves two purposes: first, to encode the synthetic history of the small molecule, and second, to positionally encode the identity of the small molecule by its location upon hybridization to an oligonucleotide array. As such, in one

embodiment, the PNA sequence identifies which monomer reaction a given solid support has experienced in the synthesis of the small molecule as well as the step in the synthesis series in which the solid support visited the monomer reaction. The PNA identifier tag can be covalently attached to the solid support or, alternatively, it can be covalently attached to the 5 target compound, *i.e.*, small molecule, through a linker group. A "monomer" of a PNA tag can include a unit of one or more PNAs that identify a particular building block used for compound synthesis. For example, a PNA monomer having a 3-base sequence "ACT" could signify an addition of a thioacetic acid monomer to a target compound. In Figure 3, for example, "x" represents the number of PNA bases in each monomer.

10 [53] "Peptide" refers to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a "polypeptide." When the amino acids are  $\alpha$ -amino acids, either the L-optical isomer or the D-optical isomer can be used. Additionally, unnatural amino acids, for example,  $\beta$ -alanine, phenylglycine and homoarginine are also included. Commonly encountered amino acids that are not gene- 15 encoded may also be used in the present invention. All of the amino acids used in the present invention may be either the D - or L -isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see*, Spatola, A. F., in **CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS**, B. Weinstein, eds., Marcel Dekker, New York, p. 267 (1983).

20 [54] "Oligonucleotides" refers to a single-stranded DNA or RNA molecule, typically prepared by synthetic means. The oligonucleotides employed in the methods of the present invention will usually be 8 to 150 nucleotides in length, preferably from 10 to 50 nucleotides, although oligonucleotides of different length may be appropriate in some circumstances. Suitable oligonucleotides may be prepared by the phosphoramidite method 25 described by Beaucage and Carruthers, *Tetr. Lett.*, 22:1859-1862 (1981), or by the triester method according to Matteucci, *et al.*, *T. Am. Chem. Soc.*, 103:3185 (1981), both incorporated herein by reference, or by other methods such as by using commercial automated oligonucleotide synthesizers.

30 [55] As used herein, the term "linking group" refers to a group that links a target compound to a solid support or a PNA identifier tag to either a solid support or a target compound. Linking groups of diverse structures are useful in practicing the present invention. Exemplary linking groups include, but are not limited to, organic functional groups (*e.g.*, -C(O)-, -NR-, -C(O)S-, -C(O)NR-, *etc.*); substituted or unsubstituted alkyl,

substituted or unsubstituted heteroalkyl and substituted or unsubstituted aryl groups each of which are, in addition to other optional substituents, homo- or hetero-disubstituted with organic functional groups, that adjoin the linker arm to, for example, the target compound and the solid support. The linking groups of the invention can include a group that is cleaved by, for example, light, heat, reduction, oxidation, hydrolysis or enzymatic action (e.g., nitrophenyl, disulfide, ester, etc.). Alternatively, the linking group can be substantially stable under a range of conditions. By providing for the use of linkers with a wide range of physicochemical characteristics, selected properties of the target compounds and their PNA identifier tags can be manipulated. Properties that are amenable to manipulation include, for example, hydrophobicity, hydrophilicity, surface-activity and the distance from the solid support of the species bound to the solid support via the linking group.

[56] The term "protecting group" or "compatible protecting group" refers to a chemical group that exhibits the following characteristics: 1) reacts selectively with the desired functionality in good yield to give a derivative that is stable to the projected reactions for which protection is desired; 2) can be selectively removed chemically and/or enzymatically from the derivatized solid support to yield the desired functionality; and 3) is removable in good yield by reagents compatible with the other functional group(s) generated in such projected reactions. Examples of protecting groups can be found in Greene, *et al.* (1991) *Protective Groups in Organic Synthesis*, 2nd Ed. (John Wiley & Sons, Inc., New York). Preferred protecting groups include, but are not limited to, acid-labile protecting groups (such as Boc or DMT); base-labile protecting groups (such as Fmoc, Fm, phosphonioethoxycarbonyl (Peoc), etc.); groups which may be removed under neutral conditions (e.g., metal ion-assisted hydrolysis), such as DBMB, allyl or alloc, 2-haloethyl; groups which may be removed using fluoride ion, such as 2-(trimethylsilyl)ethoxymethyl (SEM), 2-(trimethylsilyl)-ethyloxycarbonyl (Teoc) or 2-(trimethylsilyl)ethyl (Te) S; and groups which may be removed under mild reducing conditions (e.g., with sodium borohydride or hydrazine), such as Lev. Particularly preferred protecting groups include, but are not limited to, Fmoc, Fm, Menpoc, Nvoc, Nv, Boc, CBZ, allyl, alloc (allyloxycarbonyl), Npeoc (4-nitrophenethyloxycarbonyl), Npeom (4-nitrophenethyloxymethyloxy),  $\alpha,\alpha$ -dimethyl-3,5-dimethoxybenzyloxycarbonyl (ddz) and trityl groups. The particular removable protecting group employed is not critical to the methods of the present invention.

[57] The term "orthogonal protecting groups" refer to two or more compatible protecting groups which, in the presence of one other, can be differentially

removed or, if not differentially removed, can be differentially reprotected. In one embodiment, it may be desirable to remove all of the protecting groups in one step, such as at completion of the synthesis.

[58] The term "stereoisomer" refers to a chemical compound having the same molecular weight, chemical composition, and constitution as another, but with the atoms grouped differently. That is, certain identical chemical moieties are at different orientations in space and, therefrom, when pure, have the ability to rotate the plane of polarized light. However, some pure stereoisomers may have an optical rotation that is so slight that it is undetectable with present instrumentation. The compounds described herein 10 may have one or more asymmetrical carbon atoms and therefore include various stereoisomers. All stereoisomers are included within the scope of the invention.

[59] A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, labels suitable for use in the present invention include, for example, 15 radioactive labels (*e.g.*,  $^{32}\text{P}$ ), fluorophores (*e.g.*, fluorescein), electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, *e.g.*, by incorporating a radiolabel into the hapten or peptide, or used to detect antibodies specifically reactive with the hapten or peptide.

[60] The term "chemical library" or "array" refers to an intentionally 20 created collection of differing target compounds or molecules that can be prepared either synthetically or biosynthetically and that can be screened for biological activity in a variety of different formats (*e.g.*, libraries of soluble compounds, libraries of compounds tethered to solid supports, *etc.*). The term is also intended to refer to an intentionally created collection of stereoisomers. The library comprises at least 2 members, preferably at least 10 members, 25 more preferably at least  $10^2$  members and still more preferably at least  $10^3$  members.

---

Particularly preferred libraries comprise at least  $10^4$  members, more preferably  $10^5$  members and still more preferably at least  $10^6$  members.

[61] The term "combinatorial synthesis strategy" or "combinatorial 30 chemistry" refers to an ordered strategy for the parallel synthesis of diverse compounds by sequential addition of reagents (monomers) that leads to the generation of large chemical libraries. Thus, combinatorial chemistry refers to the systematic and repetitive, covalent connection of a set of different "monomers" of varying structures to each other to yield large arrays of diverse compounds or molecular entities.

C. Synthesis of Combinatorial Libraries Using Split-and-Pool Methodology

[62] Synthetic chemical libraries produced by combinatorial synthesis are important tools for both the chemist and the biologist. Typically, combinatorial synthesis is conducted via a multi-step synthesis to provide a library of target compounds. Each step in this synthesis involves a chemical modification of the then existing molecule formed from the previous step, wherein one can vary the choice of reagents and/or reaction conditions to provide for a variety of different target compounds. For example, such steps could include the use of different building blocks to form different compounds, the use of different inorganic or organic reagents that alter where the building blocks are added, the stereochemistry of the addition, *etc.*

[63] Many of the combinatorial approaches devised to prepare such libraries rely on solid-phase synthetic techniques and exploit the efficient split-and-pool method to assemble all possible combinations of a set of chemical building blocks. The split-and-pool method employs a pool of solid supports that contains or can be derivatized to contain reactive moieties for forming the molecules of interest tethered to the solid support. This pool is initially split and each split pool is then subjected to a first reaction that results in different modifications to each of the pools. After reaction, the pools of solid supports are combined and the pooled supports are then again split. Each split pool is subjected to a second reaction that is different for each of the pools. The process is continued until a library of target compounds is formed on the solid supports.

[64] U.S. Patent Nos. 5,708,153, 5,770,358, 6,140,493, 6,143,497 and 6,165,717, all of which have issued to Dower *et al.*, disclose the synthesis of diverse collections of oligomers (*i.e.*, peptides) using the split-and-pool methodology. As a specific example of the method disclosed therein, one may consider the synthesis of peptides three residues in length, assembled from a monomer set of three different monomers: A, B, and C. The first monomer is coupled to three different aliquots of beads, each different monomer in a different aliquot, and the beads from all the reactions are then pooled. The pool now contains approximately equal numbers of three different types of solid supports, with each type characterized by the monomer in the first residue position. The pool is mixed and redistributed to the separate monomer reaction tubes or vessels containing A, B, or C as the monomer. The second residue is coupled. Following this reaction, each tube now has beads with three different monomers in position one and the monomer contained in each particular second reaction tube in position 2. All reactions are pooled again, producing a mixture of beads each bearing one of the nine possible dimers. The pool is again distributed among the

three reaction vessels, coupled, and pooled. This process of sequential synthesis and mixing yields beads that have passed through all the possible reaction pathways, and the collection of beads displays all trimers of three amino acids ( $3^3 = 27$ ). Thus, a complete set of the trimers of A, B, and C is constructed.

5 [65] Again, the reactions employed at each stage of the synthesis can include the addition of different building blocks to the solid support, the use of different reagents and/or reaction conditions to differentially alter the existing chemical entity on the solid support, *etc.* Also combinations of different building blocks with different reagents and/or reaction conditions can also be employed.

10 [66] The split-and-pool protocol is particularly well-suited to the generation of large libraries, and the synthetic target compounds can be screened for interaction with the analyte of interest (*e.g.*, enzymes, macromolecular receptors, *etc.*) either in binding assays where the compounds remain tethered to their synthetic supports, or in soluble assays after cleavage of the compounds from the resin.

15 **D. Target Compounds**

[67] The split-and-pool method of assembling small molecules or oligomers from many types of different monomers requires that the appropriate coupling chemistry for a given set of monomer units or building blocks be used. Any set of building blocks that can be attached to one another in a step-by-step fashion can serve as the monomer set. The 20 attachment can be mediated by chemical, enzymatic, or other means, or by a combination of any of these means.

[68] The resulting small molecules or oligomers can be linear, cyclic, branched, or assume various other conformations as will be apparent to those skilled in the art. In a preferred embodiment, the small molecules or oligomers are peptides and the 25 monomers are amino acids. Suitable amino acids include, but are not limited to, L-amino acids, D-amino acids,  $\alpha$ -amino acids,  $\beta$ -amino acids and  $\omega$ -amino acids. Techniques for solid state synthesis of peptides are described, for example, in Merrifield, *J. Amer. Chem. Soc.*, 85:2149-2156 (1956). Peptide coupling chemistry is also described in *The Peptides*, Vol. 1 (eds. Gross, E., and J. Meienhofer, Academic Press, Orlando (1979)), which is incorporated 30 herein by reference.

[69] To synthesize the small molecules or oligomers, a collection of a large number of the solid supports is apportioned among a number of reaction vessels. In each reaction, a different monomer is coupled to the growing oligomer chain. The monomers may

be of any type that can be appropriately activated for chemical coupling or accepted for enzymatic coupling. Because the reactions may be contained in separate reaction vessels, even monomers with different coupling chemistries can be used to assemble the oligomers (see, *The Peptides, supra*). In a preferred embodiment, the monomer reactions are carried out 5 in parallel. After each coupling step, the solid supports on which are synthesized the oligomers of the library are pooled and mixed prior to re-allocation to the individual vessels for the next coupling step. This shuffling process produces solid supports with many oligomer sequence combinations. The sequence for any given oligomer is determined by the synthesis pathway (type and sequence of monomer reactions) for any given solid support at 10 the end of the synthesis.

[70] The length of the oligomer or the number of functional groups introduced into the molecule can vary. Typically, the number of monomers or functional groups is less than about 20. In a preferred embodiment, the number is from about 3 to about 15 and, in some embodiments, from about 6 to about 12. Protective groups known to those 15 skilled in the art can be used to prevent spurious coupling (see, *The Peptides, supra*, Vol. 3, which is incorporated herein by reference).

[71] It will be readily apparent to those of skill in the art that modifications of the split-and-pool methodology are also possible. For instance, the monomer set can be expanded or contracted from step to step or, alternatively, the monomer set could be changed 20 completely from step to step (e.g., amino acids can be used in one step, nucleosides can be used in another step, carbohydrates can be used in yet another step), provided the appropriate coupling chemistry is employed (see, Gait, *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford (1984); Friesen and Danishefsky, *J. Amer. Chem. Soc.*, 111:6656 (1989); and Paulsen, *Angew. Chem. Int. Ed. Engl.*, 25:212 (1986), all of which are incorporated 25 herein by reference).

[72] In addition, a given monomer unit can be a single monomer unit or a string of monomer units that are attached to the solid support as a single entity. For instance, a monomer unit for peptide synthesis can be, for example, a single amino acids or a larger peptide unit comprising a string of amino acids, or a combination of both. One variation is to 30 form several pools of various sequences on solid supports to be distributed among different monomer sets at certain steps of the synthesis. By this approach, one can also build oligomers of different lengths with either related or unrelated sequences, and one can fix certain monomer residues at some positions, while varying other monomer residues at other points to construct oligomer frameworks, wherein certain residues or regions are altered to

provide diversity. For instance, one may want to change only 3 to 6 amino acids in a peptide that is 6 to 12 amino acids long, keeping the remaining amino acids constant for each of the peptides synthesized. In this embodiment, the constant regions can be added as larger peptide units, whereas the variable regions can be added as single amino acids.

5        ***E. PNA Identifier Tags***

[73] Once the combinatorial library has been synthesized, the small molecule or oligomer sequence on each of the recovered solid supports must be identified. The present invention provides a method for identifying the composition and/or sequence of any of the small molecules or oligomers in the library. By tracking the synthesis pathway 10 that each small molecule or oligomer has taken, one can deduce the sequence of monomers of any small molecule or oligomer. The method of the present invention involves linking a peptido nucleic acid (PNA) identifier tag to the small molecule or oligomer or, alternatively, to the solid supports that indicates the monomer reactions and corresponding step numbers 15 that define each small molecule or oligomer in the library. After a series of synthesis steps (and concurrent PNA identifier tag additions), one "reads" the PNA identifier tag(s) associated with the small molecule or oligomer on any given solid support. In a preferred embodiment, the PNA identifier tag(s) is read by hybridizing the library of small molecules or oligomers to a spatially addressable oligonucleotide array.

[74] The PNA identifier tag can be associated with the small molecule or 20 oligomer through a variety of mechanisms, either directly, through a linking group, or through a solid support upon which the oligomer is synthesized. In the latter embodiment, one could also attach the PNA identifier tag to another solid support that, in turn, is bound to the solid support upon which the small molecule or oligomer is synthesized. In a preferred embodiment, the PNA identifier tag is associated with the small molecule or oligomer such 25 that when the small molecule or oligomer is removed from the solid support the PNA identifier tag is attached to the small molecule or oligomer, typically through a linking group. It is important to note that the PNA identifier tag does not interfere with the biological activity and/or properties of the target compound.

[75] The length of the PNA identifier tag can vary. Typically, the PNA 30 identifier tag is from about 3 to about 50 nucleotides in length. In a preferred embodiment, the PNA identifier tag is from about 6 to about 20 nucleotides in length. In another preferred embodiment, the PNA identifier tag is about 12 nucleotides in length. In certain embodiment, the PNA identifier tag can further comprise a label. Suitable labels include, but

are not limited to, fluorophores, radioactive labels, *etc.* It will be readily apparent to those of skill in the art that the label can be attached to (or immobilized on) a monomer(s) of the PNA identifier tag, a linker group that attaches the PNA identifier tag to the solid support or to the target compound or the end of the PNA identifier tag. In this latter embodiment, the PNA 5 identifier tag is capped with the label. It is also noted that in certain embodiments, the small molecule or oligomer can further comprise a label. In this embodiment, the label, such as a fluorophore or radioactive label, can be attached, *e.g.*, to a monomer of the small molecule or oligomer, either directly or through a linking group.

[76] As with the oligomer monomer units, a given monomer unit of the 10 PNA tag can be a single PNA base (*i.e.*, a single nucleotide) or a string of PNA bases (*i.e.*, a string of nucleotides that are, *e.g.*, 2, 3, 4 or 5 nucleotides in length) that are attached to the target compound or the solid support as a single entity. In a preferred embodiment, a given monomer unit of the PNA tag is a string of PNA bases that are added as a single entity. It will be readily apparent to those of skill that when only a small number of monomer units of 15 an oligomer are varied, one may need to identify only those monomers which vary among the oligomers, as when one wants to vary only a few amino acids in a peptide. For instance, one might want to change only 3 to 6 amino acids in a peptide that is 6 to 12 amino acids long, or one might want to change as few as 5 amino acids in a peptides that is 50 amino acids long. One may uniquely identify the sequence of each peptide by providing for each solid support a 20 PNA identifier tag specifying only the amino acids varied in each sequence, as will be readily appreciated by those skilled in the art. In such cases, all solid supports may remain in the same reaction vessel for the addition of common monomer units and apportioned among different reaction vessels for the addition of distinguishing monomer units.

[77] In view of the foregoing, there are several ways that the PNA can be 25 used as identifier tags. In one embodiment, the PNA can be assembled base-by-base before, during, or after the corresponding oligomer (*e.g.*, peptide) synthesis step. In one case of base-by-base synthesis, the tag for each step is a single nucleotide, or at most a few nucleotides (*i.e.*, 2 to 5). This strategy preserves the order of the steps in the linear arrangement of the PNA chain grown in parallel with the oligomer. In another embodiment, a block-by-block 30 approach is employed. In this embodiment, sets or blocks of PNAs (*e.g.*, 2, 3, 4 or 5 to 10 or more bases) are added as protected, activated blocks. Each block carries the monomer-type information, and the order of addition represents the order of the monomer addition reaction. Alternatively, the block may encode the oligomer synthesis step number as well as the monomer-type information.

[78] As noted above, the PNA identifier tags can be attached to chemically reactive groups (unmasked thiols or amines, for example) on the surface of a synthesis support that has been functionalized to allow synthesis of an oligomer and attachment or synthesis of the PNA identifier tag. Alternatively, the PNA identifier tags can be attached to chemically reactive groups on the small molecule or oligomer, typically through a linker. For instance, the PNA identifier tags can also be attached to a monomer(s) that is incorporated into the oligomer chain, or to reactive sites on linkers joining the oligomer chains to the solid support, or to reactive sites on linkers attached to the oligomer chains.

[79] In one embodiment, the solid supports will have chemically reactive groups that are protected using two different or "orthogonal" types of protecting groups. The solid supports will then be exposed to a first deprotection agent or activator, removing the first type of protecting group from, for example, the chemically reactive groups that serve as the small molecule or oligomer synthesis sites. After reaction with the first monomer, the solid supports will then be exposed to a second activator which removes the second type of protecting group, exposing, for example, the chemically reactive groups that serve as PNA identifier tag attachment sites. One or both of the activators may be in a solution that is contacted with the supports.

[80] In another embodiment, the linker joining the oligomer and the solid support may have chemically reactive groups protected by the second type of protecting group. After reaction with the first monomer, the solid support bearing the linker and the "growing" oligomer will be exposed to a second activator which removes the second type of protecting group exposing the site that attaches the identifier tag directly to the linker, rather than attachment directly to the solid support.

[81] As noted above, the invention can also be carried out in a mode in which the PNA identifier tag is attached directly (or through a linker) to the oligomer being synthesized. Again, in this embodiment, when the small molecule or oligomer is removed from the solid support, the PNA identifier tag remains attached to the small molecule or oligomer. The size and composition of the library will be determined by the number of coupling steps and the monomers used during the synthesis. Those of skill in the art recognize that either the monomer of the PNA identifier tag or the monomer of the oligomer may be coupled first, in either embodiment.

[82] In addition to encoding the synthetic history of the small molecule or oligomer, the PNA identifier tag of the present invention also serves to positionally encode the identity of the small molecule by its location upon hybridization to an oligonucleotide

array. The sequences of the PNA identifier tags are initially selected such that they are capable of hybridizing to known sequences on the oligonucleotide array. Methods of making arrays of oligonucleotides are known to those of skill in the art (see, e.g., U.S. Patent No. 5,143,854, the teachings of which are incorporated herein by reference). Moreover, arrays of 5 oligonucleotides are available from a number of commercial sources, such as Affymetrix (Santa Clara, California). In a preferred embodiment, a GenFlex™ tag array, which is commercially available from Affymetrix, is employed (arrays of this type are currently available at a density of 400,000 features/cm<sup>2</sup>; the sequences of the chip's probes are available from Affymetrix). In the GenFlex™ tag array, the oligonucleotides are about 20 10 nucleotides in length and, thus, the sequences of the PNA identifier tag can be selected to hybridize to the full-length sequences of the oligonucleotide probes or to a portion of the sequences of the oligonucleotide probes. In a preferred embodiment, the PNA sequences are selected to hybridize to the terminal 12 residues of the 20 mer probes of a GenFlex™ tag array.

15 [83] Once the PNA identifier tags have hybridized to the array of oligonucleotides, they can be detected using a variety of different means. For instance, if the PNA identifier tag is labeled with a fluorophore, the array or chip can be scanned for fluorescence. The location of the fluorescence reveals the sequence of the PNA identifier tag and, in turn, the structure of the library member. Similarly, if the PNA tag is labeled with a 20 radioactive label, the location of the radioactivity reveals the sequence of the PNA identifier tag and, in turn, the structure of the library member. Other labeling and detection systems suitable for use in the methods of the present invention will be readily apparent to those of skill in the art.

25 [84] As noted, in certain embodiments, the PNA identifier tag can further comprise a label. Suitable labels include, but are not limited to, fluorophores, radioactive labels, etc. It is also noted that in certain embodiments, the small molecule or oligomer can further comprise a label. In this embodiment, the label, such as a fluorophore or radioactive label, can be attached, e.g., to a monomer of the small molecule or oligomer, either directly or through a linking group.

30 [85] Some of the features of the PNA identifier tags of the present invention include, but are not limited to, one or more of the following: (a) the PNA identifier tag does not interfere with the biological activity or properties of the target compound; (b) detection limits for the PNA identifier tag are very low; (c) reaction conditions for attaching the PNA identifier tag to the solid support or the target compound are mild enough not to affect the

synthesis of the target compound; (d) the PNA identifier tag is stable under various reaction conditions; and (e) the PNA identifier tag can be easily synthesized in large quantity and in large scale.

***F. Solid Supports/Linkers***

5 [86] The chemical or enzymatic synthesis of the small molecule or oligomer libraries of the present invention typically takes place on solid supports. The term "solid support," as used herein, embraces a substrate with appropriate sites for oligomer synthesis and, in some embodiments, PNA identifier tag attachment and/or synthesis. There are various solid supports useful in the preparation of the synthetic oligomer libraries of the 10 present invention. In fact, synthesis on solid supports, "solid-phase synthesis," is of recognized utility in the synthesis of small molecules, oligomeric compounds and polymers. A diverse array of solid supports bearing useful reactive groups are known in the art (see, for example, Burgess, ed., SOLID-PHASE ORGANIC SYNTHESIS, John Wiley and Sons (2000); and Chan and White, eds., FMOC SOLID PHASE PEPTIDE SYNTHESIS: A PRACTICAL APPROACH 15 (The Practical Approach Series), Oxford University Press (2000)). Solid supports include substantially any oligomeric or polymeric material upon which a selected synthesis can be performed, and the materials and methods of the present invention are not limited by the identity of the material serving as the solid support.

20 [87] With enough solid supports and efficient coupling, one can, if desired, generate complete sets of certain oligomers. In general, the size of the solid support is in the range of 1 nm to 100  $\mu$ m, but a larger solid support of up to 1 mm in size can be used. To improve washing efficiencies, solid supports less porous than typical peptide synthesis resins are preferable. As such, in a preferred embodiment, the solid support is nonporous. Solid supports can be of any shape, although they will preferably be roughly spherical (e.g., beads, 25 particles, etc.). The supports need not necessarily be homogenous in size, shape, or composition; although the supports usually and preferably will be uniform. In some embodiments, supports that are very uniform in size and shape may be particularly preferred. In another embodiment, however, two or more distinctly different populations of solid supports may be used for certain purposes.

30 [88] Solid supports can consist of many different materials, limited primarily by capacity for derivatization to attach any of a number of chemically reactive groups and compatibility with the chemistry of small molecule or oligomer synthesis and PNA identifier tag synthesis and attachment. Suitable solid support materials include, but are

not limited to, glass supports, latex supports, silicon dioxide supports containing Si-OH groups, polystyrene supports, polyacrylamide supports, polyethyleneglycol supports and the like, gold or other colloidal metal particles, and other materials known to those skilled in the art. Preferred solid supports include, but are not limited to, Rink Amide MBHA resin, *p*-benzyloxybenzyl alcohol resin (Wang), 4-hydroxymethyl benzoic acid resin, 4-sulfamylbenzoyl resin, and the like. Except as otherwise noted, the chemically reactive groups with which such solid supports may be derivatized are those commonly used for solid state synthesis of the respective oligomer and thus will be well known to those skilled in the art.

[89] One of the monomers employed in the synthesis is or becomes covalently attached to the solid support such that the target compound resulting from the synthetic scheme employed is covalently attached to the support. Preferably, such covalent attachment is through a linking group or, interchangeably, a linking arm. Suitable linking groups are well known in the art and include, but are not limited to, conventional linking groups such as those comprising esters, amides, carbamates, ethers, thio ethers, ureas, amines and the like.

[90] The linking group can be cleavable or non-cleavable. "Cleavable linking groups" refer to linking groups, wherein at least one of the covalent bonds of the linking group that attaches, *e.g.*, the target compound to the solid support can be readily broken by specific chemical reactions, thereby providing for target compounds free of the solid support ("soluble compounds"). The chemical reactions employed to break the covalent bond of the linking arm are selected so as to be specific for bond breakage, thereby preventing unintended reactions occurring elsewhere on the target compound or the PNA identifier tag. That is to say, the cleavable linking group is selected relative to the synthesis of the compounds to be formed on the solid support (*i.e.*, target compounds or PNA identifier tags) so as to prevent premature cleavage from the solid support as well as not to interfere with any of the procedures employed during compound synthesis on the support.

[91] Suitable cleavable linking arms are well known in the art. For instance, a cleavable Sasrin resin comprising polystyrene beads and a cleavable linking arm, which linking arm is cleaved by strong acidic conditions such as trifluoroacetic acid, can be used. Similarly, cleavable TENTAGEL AC, TENTAGEL PHB and TENTAGEL RAM can be used. Reversible covalent cleavable linkages can also be used to attach the target compounds to the solid supports. Examples of suitable reversible chemical linkages include, but are not limited to, (1) a sulfoester linkage provided by, *e.g.*, a thiolated tagged-molecule

and a N-hydroxy-succinimidyl support, which linkage can be controlled by adjustment of the ammonium hydroxide concentration; (2) a benzylhydryl or benzylamide linkage provided by, *e.g.*, a Knorr linker, which linkage can be controlled by adjustment of acid concentration; (3) a disulfide linkage provided by, *e.g.*, a thiolated tagged-molecule and a 2-pyridyl disulfide support (*e.g.*, thiolsepharose from Sigma), which linkage can be controlled by adjustment of the DTT (dithiothreitol) concentration; and (4) linkers which can be cleaved with a transition metal (*e.g.*, HYCRAM).

[92] The linker may be attached between the PNA identifier tag and/or the small molecule or oligomer and the support via a non-reversible covalent cleavable linkage.

10 For example, linkers which can be cleaved photolytically can be used. Preferred photocleavable linkers of the invention include, but are not limited to, 6-nitro-veratryloxycarbonyl (NVOC) and other NVOC related linker compounds (*see*, PCT Patent Publication Nos. WO 90/15070 and WO 92/10092); the ortho-nitrobenzyl-based linker described by Rich (*see*, Rich and Gurwara, *J. Am. Chem. Soc.*, 97:1575-1579 (1975); and

15 Barany and Albericio, *J. Am. Chem. Soc.*, 107: 4936-4942 (1985)); and the phenacyl based linker disclosed by Wang. (*see*, Wang, *J. Org. Chem.*, 41:3258 (1976); and Bell and Mutter, *Chimia*, 39:10 (1985)).

#### ***G. Screening of Libraries***

[93] Once prepared, the library of target compounds can be subsequently screened/assayed for biological activity or other properties, either on the solid support or after the library of compounds has been removed from the solid support. It will be readily apparent to those of skill in the art that the library of compounds can be screened/assayed for biological activity or other properties using standard assays known to and used by those of skill in the art. Properties that can be screened for include, but are not limited to, the following: biological activities, binding affinities, biological properties, pharmacological properties, oral bioavailabilities, circulatory half-lives, agonist activities, antagonist activities, solubilities, *etc.* The library of target compounds can be screened for useful properties sequentially or in parallel. Once identified, the library compounds having useful properties can be prepared on a large-scale. Methods of screening libraries are described in,

25 COMBINATORIAL LIBRARIES: SYNTHESIS, SCREENING, AND APPLICATION POTENTIAL, Cortese, R., Ed., Walter de Gruyter, Berlin, 1996, pp. 159-174, which is incorporated herein by reference.

[94] One example of the use of the method of the invention is shown in Figure 4. A library of interest is synthesized by traditional split-and-pool strategy. For the purpose of this invention, the tagging step involves coupling of the appropriate nucleotides which will denote the building block used in that step and will be ultimately be used to 5 localize the compound on a spatially addressable array. In the final step, the whole library is cleaved and obtained as a single mixture of compounds.

[95] The initial assay may be carried out in solution prior to, or after, arraying through hybridization. It is often preferable to perform the incubation of a library with the target in solution in order to avoid nonspecific interaction of the target with the array 10 surface. For example, when screening a library of small molecule inhibitors against a set of enzymes, the library members are preferably incubated with the enzymes prior to hybridization of the library member tags to the support.

[96] PNA-encoded libraries of protein ligands can be screened against several targets simultaneously by incubating the library with the various targets containing 15 different fluorophores. Upon hybridization of the mixture to an oligonucleotide array, fluorescent detection reveals the identity and selectivity of library members that bind a target. An example of this approach is shown in Figure 5A, in which a library of small molecule-tag adducts is screened against a small library of enzymes which are all labeled with a different fluorescent tag. After incubation of the library with the enzymes, the library is hybridized to 20 the DNA chip and scanned for fluorescence. This type of assay is well suited for screens involving the selective inhibition of a particular enzyme within a family of isozymes since the selectivity of a particular hit is immediately established.

[97] Although attractive for drug discovery, this strategy does not always lend itself to profiling biological samples since it is difficult to label uniformly all the 25 proteins. Conversely, the PNA-small molecule conjugate can be synthesized with a fluorophore (Figure 5B). For example, a fluorescent tag can be attached to the N-terminus of the PNA tag. Proteins or other macromolecules from lysed cells, tissues, other biological samples or industrial samples, or from a collection of, for example, enzymes or receptors, are incubated with the PNA-encoded library of interest, after which the mixture is subjected to 30 fractionation to separate library members that are bound to a macromolecule from members that are not. For example, after incubation with a sample of interest, the PNA-small molecule conjugate bound to a macromolecule can be separated from the unbound PNA-small molecule conjugate by, for example, size exclusion chromatography. Preferably, the rate of

dissociation between the small molecule and its target is slow relative to the time of size exclusion separation.

[98] In some embodiments, suitable "warhead" can be used such that a macromolecule becomes covalently attached to the library member to which the 5 macromolecule binds. Examples of suitable warheads for proteases include, but are not limited to, those shown in Figure 19. These warheads are suitable for, for example, serine, cysteine, threonine, aspartyl, and metallo-proteases.

[99] The high molecular weight material is then incubated on a DNA chip such that the tags hybridize to the chip. The chip is then scanned for fluorescence. The 10 location of fluorescence reveals the structure of the library member that bound to a macromolecule. Thus, in this example, the identity of the macromolecule is not known, but the structure of the molecule that binds to it is known. The structure of the macromolecule can then be identified by methods known to those of skill in the art. For example, one can use mass spectroscopy directly on the DNA array, or the macromolecule can be isolated in larger 15 amounts by, for example, affinity chromatography using the substrate to which it bound on the chip, for more traditional characterization.

[100] This method is useful not only for the discovery of small molecule ligands, but also for proteomic profiling and diagnostics. As shown in Figure 6, hybridization of the high molecular weight fraction to a chip reveals the identity of the small 20 molecules bound to macromolecules, thereby generating a profile of protein function. The correlation between profiles and phenotypes can be rapidly assessed in the biological system using the small molecules identified in the profile while their molecular target(s) can be determined by affinity chromatography. For example, a library of kinase inhibitors can be used to compare tissue samples such as a carcinoma and its healthy counterpart, thereby 25 revealing conspicuously over-abundant or absent kinases in the carcinoma tissue. Likewise a library of mechanism-based inhibitors, such as cysteine inhibitors, can be used to measure the activity of all cysteine proteases in a tissue sample.

[101] As noted above, several possible options are available by which to screen libraries of interest using the invention described herein. Once hybridized to the array, 30 the outcome of the assay can be detected by, for example, fluorescence measurement, whereby either the oligonucleotide tagged library member is fluorescently labeled or the analyte such as the enzyme in the previous example is fluorescently labeled. Other types of detection such as radioactive labeling using a chip coated with a scintillating material, by

surface resonance spectroscopy, atomic force microscopy, and other detection methods known to those of skill in the art are also suitable.

#### ***H. Labels***

[102] As noted above, depending on the screening assay employed, the PNA identifier tag, the target compound and/or the analyte or ligand of interest can be labeled. The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere the assay being carried out or with the specific binding of the PNA identifier tag to the oligonucleotide in the oligonucleotide array. The detectable group can be any material having a detectable physical or chemical property.

10 Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means.

[103] Examples of labels suitable for use in the present invention include, but are not limited to, fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^{32}\text{P}$ ), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (*e.g.*, polystyrene, polypropylene, latex, *etc.*).

[104] The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels can be used, with the choice of label depending on sensitivity required, ease of conjugation with the desired component of the assay (*e.g.*, PNA identifier tag), stability requirements, available instrumentation, and disposal provisions. Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (*e.g.*, biotin) is covalently bound to the molecule. The ligand then binds to another molecules (*e.g.*, streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound.

[105] The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes suitable for use as labels include, but are not limited to, hydrolases, particularly phosphatases, esterases and glycosidases, or oxidases, particularly peroxidases. Fluorescent compounds, *i.e.*, fluorophores, suitable for use as labels include, but are not limited to, fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, *etc.* Further examples of

suitable fluorophores include, but are not limited to, eosin, TRITC-amine, quinine, fluorescein W, acridine yellow, lissamine rhodamine, B sulfonyl chloride erythroscein, ruthenium (tris, bipyridinium), Texas Red, nicotinamide adenine dinucleotide, flavin adenine dinucleotide, *etc.* Chemiluminescent compounds suitable for use as labels include, but are 5 not limited to, luciferin and 2,3-dihydrophthalazinediones, *e.g.*, luminol. For a review of various labeling or signal producing systems that can be used in the methods of the present invention, see U.S. Patent No. 4,391,904.

[106] Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a 10 scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing 15 the appropriate substrates for the enzyme and detecting the resulting reaction product. Colorimetric or chemiluminescent labels may be detected simply by observing the color associated with the label. Other labeling and detection systems suitable for use in the methods of the present invention will be readily apparent to those of skill in the art.

#### *I. Other Features of the Methods of the Present Invention*

[107] As noted above, in one embodiment, the present invention provides a 20 method for preparing a library of diverse compounds, each of the compounds being produced by the step-by-step assembly of building blocks, the method comprising the steps of: (a) apportioning solid supports among a plurality of reaction vessels; and (b) in each reaction vessel of the plurality of reaction vessels, exposing the solid supports to a first building block 25 of a first compound and to a first monomer of a first peptido-nucleic acid (PNA) identifier tag under conditions suitable for immobilization of the first building block and the first monomer, wherein the first building block present in one reaction vessel is different from the first building block present in the other reaction vessels, wherein the first building block of the first compound is capable of being covalently coupled to a second building block and 30 wherein the first monomer of the PNA identifier tag is capable of being covalently coupled to a second monomer.

[108] In a preferred embodiment, any additional reactive groups of the first building block of the first compound or any additional reactive groups of the first monomer

of the first PNA identifier tag capable of interfering with subsequent couplings are suitable protected prior to subsequent couplings. In another preferred embodiment, the first monomer of the first PNA identifier tag does not interfere with the coupling of the first building block of the first compound to the second building block of the first compound. Alternatively, in a 5 preferred embodiment, the first monomer of the first compound does not interfere with the coupling of the first monomer of the first PNA identifier tag to the second monomer of the first PNA identifier tag. In another preferred embodiment, the first monomer of the first PNA identifier tag identifies the first building block of the first compound. In another preferred embodiment, the PNA identifier tag does not contribute to the activity or properties (e.g., 10 binding characteristics) of the target compound. In another preferred embodiment, the PNA identifier tag can be detected and identified, such as by hybridization to an oligonucleotide in an oligonucleotide array. In another preferred embodiment, reactive groups of the building blocks of the target compounds and the reactive groups of the monomers of the PNA identifier tags are independently selected and include, but are not limited to, amino groups, 15 hydroxyl groups, carboxyl groups and phosphate groups. The foregoing features of the methods of the present invention are intended to be illustrative and not exhaustive. Other features, embodiments and advantages of the methods of the present invention will be readily apparent to those of skill in the art upon reading this disclosure.

## EXAMPLES

20 [109] The following examples are offered to illustrate, but not to limit the present invention.

### Example 1

#### Split and Pool Synthesis of a PNA-encoded Combinatorial Library of Potential Tyrosine Kinase Inhibitors

25 [110] This Example describes a scheme for the split-and-pool synthesis of a PNA-encoded combinatorial library. This scheme (Scheme 1), which is shown in Figure 7, illustrates the synthesis of a library of potential tyrosine kinase inhibitors, which serves as a representative example of the types of libraries that can be synthesized and screened using the methods of the present invention.

30 [111] It has been demonstrated that substitution of a tyrosyl residue for a tetrafluorotyrosyl residue generates a competitive inhibitor of tyrosine kinase (Yuan, *et al.*, *J. Biol. Chem.*, 265:16205-16209 (1990)). Screening of this library as described herein is useful not only to discover inhibitors of tyrosine kinase on a proteome-wide scale, but also for

therapeutic target discovery and validation since it also provides a profile of tyrosine kinases. For instance, comparison of a screen using the crude cell extracts from a cancer cell to its respective healthy cell extracts can reveal an overabundant kinase as a potentially new therapeutic target. Since this kinase will be identified based on an inhibitor, the inhibitor may 5 be used directly in a whole cell assay to validate the therapeutic target.

#### Experimental Procedures

[112] *General procedure for amino acid coupling.* The resin was suspended in DMF (10 mL/g) and the Fmoc protected acid was added (4.0 eq., standard acid labile protecting groups were used for side chain heteroatoms, all amino acids were purchased from 10 NovaBiochem) followed by HOBr (4.0 eq.). The reaction was agitated on a wrist shaker for 4 hr after which the resin was poured in a glass filtered funnel and washed with DMF, MeOH, DMF, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O (each washing was performed with 20 mL/g of solvent).

[113] *General procedure for Fmoc deprotection.* The resin was suspended 15 in DMF (7 mL/g) and piperidine (7 mL/g) was added. The reaction was agitated on a wrist shaker for 1 hr, venting the reaction at 2, 5, 15, 30 min. The resin was then poured into a glass filtered funnel and washed with DMF, MeOH, DMF, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O (each washing was performed with 20 mL/g of solvent).

[114] *General procedure for Alloc deprotection.* The resin was suspended in 20 wet CH<sub>2</sub>Cl<sub>2</sub> (12 mL/g) and Pd(PPh<sub>3</sub>)<sub>4</sub> (0.1 eq.) was added followed by Bu<sub>3</sub>SnH (4.0 eq.). The reaction was agitated on a wrist shaker for 2 hr, venting the reaction at 2, 5, 15, 60 min. The resin was then poured into a glass filtered funnel and washed with CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O (each washing was performed with 20 mL/g of solvent).

[115] *General procedure for PNA coupling.* The resin was suspended in 25 DMF (10 mL/g) and the Alloc protected acid 6 was added (4.0 eq., Boc protecting groups were used on the nucleotide heterocycle) followed by HOBr (4.0 eq.). The reaction was agitated on a wrist shaker for 4 hr after which the resin was poured into a glass filtered funnel and washed with DMF, MeOH, DMF, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O (each washing 30 was performed with 20 mL/g of solvent).

[116] *Preparation of monoprotected bis amino resin 3.* Resin 1 (1.2 mmol/g, NovaBiochem) was suspended in DMF (10 mL/g) and triethylamine was added (3.0 eq.) followed by the anhydride 2 (2.5 eq.). The reaction was agitated for 6 hr on a wrist shaker

after which the resin was poured into a glass filtered funnel and washed with DMF, MeOH, DMF, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, MeOH, CH<sub>2</sub>Cl<sub>2</sub>, Et<sub>2</sub>O (each washing was performed with 20 mL/g of solvent).

[117] *Synthesis of Combinatorial Library.* The polymer bound amine resin 3 was split into 20 equal portion and coupled with the first amino acid 4 (20 natural amino acids) according to the general procedure. Each pool was then subjected to Alloc deprotection according to the general procedure. The structure of the amino acid used in each pool was then encoded with three rounds of nucleotide 6 coupling/deprotection (the natural encoding scheme was used). The 20 pools of resins were then repooled, thoroughly mixed and split again into 20 portion of a second round of amino acid coupling/encoding. After repooling the resin, the whole batch was subjected to a coupling with Fmoc-protected tetrafluorotyrosine. The whole batch was then subjected to another two rounds of peptide coupling/encoding using a Boc-protected amino acid rather than an Fmoc-protected amino acid in the last round to afford the polymer bound library 8. Alloc deprotection of the whole resin followed by coupling to the fluorophore Alexa 350 under the recommended protocol furnished 10 which was cleaved and fully deprotected in a single treatment with 50% TFA in CH<sub>2</sub>Cl<sub>2</sub> (10 mL/g) for 1 h. The library was concentrated and dried under high vacuum for 24h.

#### Example 2

20 Split and Pool Synthesis of a PNA-encoded Combinatorial Library  
of Potential Protease Inhibitors

[118] This Example describes the application of the PNA-encoding methodology to on mechanism-based cysteine protease inhibitors that contain an acrylamide functionality (see, Kong, *et al.*, *J. Med. Chem.*, 41:2579 (1998); Caulfield, *et al.*, *J. Combi. Chem.*, 2:600 (2000); and Walsh, *Tetrahedron*, 38:871 (1982)). Preliminary studies to determine the optimal length of PNA indicated that 12mers have good hybridization properties and allow ample sequence variation to encode very large libraries. The synthesis was carried out on acid-labile Rink resin with mutually compatible Fmoc and Alloc protecting groups for the inhibitor and PNA synthesis, respectively; the side chains and bases were protected with acid labile groups (Figure 10). All of the compounds synthesized exhibited satisfactory analytical and functional characteristics.

[119] The design of inhibitors was based on the information gathered from a previously developed method to rapidly assess the substrate specificity of proteases (see, Harris, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 97:7754 (2000); and Harris, *et al.*, P. Alper, J. Li,

M. Rechsteiner, B. J. Backes, *submitted*). A comparison of the activity of compounds 1 and 2 (Figure 10) against cathepsin C reveals that the PNA tag does not significantly affect the activity or selectivity of compounds 1 and 2 for cathepsin C relative to cathepsin L (Table 1). An additional PEG spacer was included in the library synthesis to insure good water 5 solubility.

Table 1

	cathepsin C		cathepsin L	
	IC <sub>50</sub> (μM)	k <sub>inact</sub> /K <sub>i</sub> (M <sup>-1</sup> S <sup>-1</sup> )	IC <sub>50</sub> (μM)	k <sub>inact</sub> /K <sub>i</sub> (M <sup>-1</sup> S <sup>-1</sup> )
compound 1	17.6	40	>2 000 μM	NA
compound 2	14.1	70	>2 000 μM	NA

[120] A series of compounds designed to inhibit cathepsin S, L, H, B, C and calpain were synthesized (Figure 11). The PNA sequences were selected to hybridize to the 10 terminal 12 residues of the 20 mer probes of a GenFlex™ tag array (arrays of this type are currently available at a density of 400,000 features/cm<sup>2</sup>; the sequences of the chip's probes are available from Affymetrix). The PNA tags only hybridize to a portion of the array probe and it was expected that each probe would have different hybridization properties.

[121] Hybridization of a mixture of the 6 probes (45 pmol of each in 150 mL) afforded the results shown in Figure 12, panel A. The difference in intensity of each array feature reflects the differences in melting temperature of the individual probes. Importantly, despite such differences in melting temperature, 30% changes in probe concentration were reliably detected. An equimolar mixture of the six compounds (3-8, 28 pmol) was incubated with commercially available purified cathepsin C (110 mg in 20 mL 20 buffer (100 mM NaOAc, pH 5.5; 100 mM NaCl; 1.0 mM EDTA, 0.01% Brij-35; 2.0 mM DTT) for 2 hours at 23°C, passed through a size exclusion column (BioRad, Bio-Sil, SEC 125-5) to remove material below 10 kDa and hybridized to a GenFlex™ tag array. As shown 25 in Figure 12, panel C, hybridization afforded the expected signal for the probe corresponding to cathepsin C, while a control lacking cathepsin C gave no signal (Figure 12, panel B). The same experiment was performed with cathepsin L (6 mg in 20 mL) using 10 fold less protein. Direct detection of the fluorescein gave a weak signal, but this signal could be amplified 30 using an anti-fluorescein goat Ab followed by a biotinylated anti-goat Ab and phycoerythrin labeled streptavidin (Figure 12, panel D).

[122] These results show that the proposed size exclusion separation is effective to separate the bound PNA-ligand conjugates from the unbound ones, that PNA is

efficient for positional encoding and that small molecule-PNA conjugates can be used to probe protein function in a microarray format.

### Example 3

#### Split and Pool Synthesis of a PNA-encoded Combinatorial Library

##### 5 of Potential Cysteine Protease Inhibitors

[123] This Example is directed to cysteine proteases. An acrylate moiety (Figure 14) was selected as the mechanism-based "war-head" based on its chemoselectivity for nucleophilic thiols (see, Dragovich, *et al.*, *J. Med. Chem.*, 41:2806-2818 (1998); and Leung, *et al.*, *J. Med. Chem.*, 43:305-341 (2000)). To examine the ability of this method to 10 quantitatively monitor changes in active-enzyme amounts and to do so in complex physiologically relevant samples, cytotoxic lymphocyte mediated cell death was demonstrated. The results demonstrate that the methods of the present invention can be used to monitor proteolytic activities in complex biological processes that are not regulated at the protein synthesis level.

##### 15 Experimental Procedures

[124] *Preparation of compounds 1-7.* Unless otherwise indicated, the chemicals were purchased from Aldrich, IL and reactions were performed at room temperature. The Fmoc protected amino acrylic acids were prepared from the corresponding Fmoc protected amino acid (NovaBiochem, CA) via a four steps sequence. Esterification of 20 the amino acid with ethane thiol and WSC (NovaBiochem, CA) in dichloromethane afforded the thio ester which was reduced to the corresponding aldehyde with 10% palladium on charcoal and triethylsilane in dichloromethane (see, Fukuyama, *et al.*, *J. Am. Chem. Soc.*, 112:7050-7051 (1990)). The aldehyde was condensed with Allyl (triphenylphosphor-snlylidene)acetate in toluene at 80°C to obtain the allyl protected trans acrylate. The 25 geometry of the olefin was verified by NMR ( $J=11.5$  Hz). The allyl group was removed using palladium tetrakis (Strem, NH) and tributyltin hydride in dichloromethane. The peptide PNA conjugates were synthesized on Rink Amide MBHA resin (NovaBiochem, CA) using Fmoc-Lys(Mtt)OH as the first and branchpoint residue. The PNA synthesis was carried out 30 using an Applied BioSystem Expedite synthesizer according to the manufacturer's recommendations.

[125] *Enzyme and apoptotic lysate preparation.* Human caspase-3 was cloned, expressed, and purified by methods previously described by Zhou and Salvesen, *et al.* (see, Zhou, Q., *et al.*, *J. Biol. Chem.*, 272:7797-7800 (1997)). Human granzyme B was

cloned and expressed in *Pichia pastoris* utilizing methods previously described (see, Harris, *et al.*, *J. Biol. Chem.*, 273:27364-27373 (1998)), with the exception that a C-terminal 6xHis tag was incorporated to facilitate purification on Ni(II) resin. The Jurkat cytosolic cell lysates were prepared by lysing 10 x 10<sup>6</sup> cells in a buffer consisting of 10 mM Hepes pH 7.4, 130 mM NaCl, and 1% Triton X-100. The soluble cytosolic fraction was separated from the insoluble membrane and nuclear fraction through centrifugation at 12 k rpm for 10 minutes. The soluble cytosolic lysate was adjusted to 1 mg/mL by the addition of PBS and 5 mM DTT. To make the granzyme B-activated apoptotic lysate, recombinant granzyme B was added to a final concentration of 0.1 nM and incubated for 30 minutes or until caspase activity reached a plateau, as monitored by Ac-DEVD-acc fluorescence (see, Harris, *et al.*, *Proc. Natl. Acad. Sci. USA*, 97:7754-7759 (2000)).

[126] *Incubation of mechanism-based probes (1-7) with enzyme and lysate samples.* Compounds 1-7 were incubated at 1.0  $\mu$ M in 20  $\mu$ L with purified caspase-3, purified granzyme B, cytosolic lysate from Jurkat cells, or granzyme B activated apoptotic Jurkat lysates for 2 h in PBS pH 7.4 supplemented with 5 mM DTT. The sample was then loaded on an ultrafree 30 kDa molecular weight cutoff filter (Millipore, MA) and washed with 1x PBS buffer (3 x 500  $\mu$ L). The volume of the sample retained in the 30 kDa filter was then adjusted to 200  $\mu$ L with PBS and fluorescein-conjugated DNA control probes were added to the sample. The sample mixture was then added to a GenFlex<sup>TM</sup> tag array (Affymetrix) and was visualized after a 6 hour incubation.

[127] *Capture of protein functionally interacting with probe 6a.* Granzyme B activated apoptotic jurkat lysates (prepared as described above) were incubated with compound 6b, for 1 hour. Ultralink immobilized monomeric avidin resin was then added to the sample and incubated at room temperature for 1 hour. The resin was then washed with 10 x resin volume of PBS and captured proteins were eluted with 5 mM biotin.

[128] *Identification of protein functionally interacting with probe by mass spectrometry.* Captured proteins were denatured with 8M urea in 100 mM ammonium carbonate and then reduced by adding dithiothreitol to a final concentration of 10 mM and incubating for 45 min at 50°C. Iodoacetamide was added to a final concentration of 30 mM and the resulting solution was allowed to stand at room temperature for 45 min. Proteins were digested with sequencing grade modified trypsin (ProMega, Madison, WI) according to the manufacturer's instructions. Tryptic peptides were analyzed by nanoflow RP-HPLC/ $\mu$ ESI/MS on an LCQ quadrupole ion trap mass spectrometer (ThermoFinnigan, San

Jose, CA). Briefly, 10 pmol of tryptic peptides were loaded onto a microcapillary column (360  $\mu$ m O.D. x 75  $\mu$ m I.D. fused silica) packed with 6 cm 5-20  $\mu$ m C18 particles (Waters, Milford, MA). This column was connected to an analytical column (360  $\mu$ m O.D. x 50  $\mu$ m I.D. fused silica packed with 8 cm 5  $\mu$ m C18) with an integrated ESI emitter tip. The 5 construction of this type of column and its use in ESI-MS has been described previously (see, Martin, et al., *Analytical Chem.*, 72:4266-4274 (2000)). Peptides were eluted into the mass spectrometer with an HPLC gradient consisting of 0-70% B in 20 minutes (A = 0.1 M acetic acid in water, B = acetonitrile with 0.1 M acetic acid). The mass spectrometer was programmed to record continuous cycles of MS scans (*m/z* 300-2000) followed by MS/MS 10 scans of the three most abundant ions in each MS scan (collision energy 35%). MS/MS spectra were matched to peptide sequences in NCBI's non-redundant protein database (ncbi.nlm.gov/blast/db/nr.Z) using the SEQUEST algorithm (see, Eng, J., *J. Am. Soc. Mass. Spec.*, 5:976-989 (1994)).

[129] *Inhibition of the caspase executed apoptotic phenotype.* Jurkat 15 cytosolic lysate was incubated with and without 1.0  $\mu$ M compound 6c for 10 minutes. Granzyme B was then added to the lysates and caspase activity was monitored by Ac-DEVD-acc fluorescence. Aliquots were removed before the addition of granzyme B and 1, 5, 10, and 20 minutes after the addition of granzyme B. Upon removal of the aliquots, the reactions were quenched by addition of gel-loading buffer and heat denaturation. Controls for stability 20 of proteins in the lysate without the addition of granzyme B and in the presence of the inhibitor were also collected. Samples were run on 10-20% SDS-PAGE and transferred to nitrocellulose and probed with anti-caspase-3 antibody to the N-terminus of the P17 subunit (Sigma) and anti-DFF45 C-terminus antibody (Sigma). Whole Jurkat cells ( $5 \times 10^5$ ) were incubated for 12 hours with and without 10 ng/mL Anti-fas antibody, CH-11 (Kamiya 25 Biomedical Co., Seattle, WA) and with and without 1  $\mu$ M Cbz-Asp(OMe)-Glu(OMe)-Val-Asp(OMe)-FMK (Enzyme Systems Products, Livermore, CA). Cells were prepared for FACS by staining with Annexin V conjugated to enhanced green fluorescent protein (MBL, Naka-Ku Nagoya, Japan) and propidium iodide. The stained cells were then analyzed by flow cytometry.

30 **Results and Discussion**

[130] *Evaluation of the sensitivity and linearity of the method.* Peptide acrylates (AcrXxx, Fig 2) covalently and irreversibly modify cysteine proteases through Michael addition to the active site cysteine. Specificity of the acrylates for particular

proteases can be achieved through modification of the peptide moiety. It is important to note that such acrylates are remarkably stable to non-activated thiols such as dithiothreitol (DTT), glutathione or dithiothreitol, thiols that are found in biological samples and buffers. Thus, combinatorial peptide libraries containing the acrylate "war-head" should allow for the 5 discovery of novel activities or even as-yet-unidentified enzymes. In addition, peptide-acrylates specifically targeting particular proteases can be designed by utilizing the optimal peptide sequence determined from substrate specificity libraries (see, Harris, *et al.*, *Proc. Natl. Acad. Sci. USA*, 97:7754-7759 (2000); and Harris, *et al.*, *Chem. Biol.*, 8:1131-1141 (2001)). For the purpose of this study, several peptide acrylates were designed to selectively 10 target several cellular proteases, members of the cathepsin family and the caspase family.

[131] To demonstrate that PNA-encoded activity probes can quantitatively determine the differences in active protein concentrations, PNA-inhibitor adducts 1-7 (Figure 12), including the caspase-3 inhibitor (6a), were incubated with purified caspase-3 at multiple discrete concentrations ranging from 10 to 500 nM. The incubation was carried out in 20  $\mu$ L 15 of PBS buffer with 5 mM DTT using 20 pmol of each probes (1  $\mu$ M) for 2 h. The unbound probes were removed by a simple filtration through a 30 kDa molecular weight cutoff filter. The retained sample was then hybridized to a GenFlex<sup>TM</sup> oligonucleotide microarray (www.affymetrix.com) and directly imaged by fluorescein fluorescence. As shown in Figure 15, a good correlation was observed (standard deviation of less than 10%) between the 20 concentration of active caspase-3 in the assay solution and the fluorescence of the corresponding site on the microarray chip. It is important to note that the intensity and contrast of the images shown in Figure 15 was standardized for comparison purposes. The probe corresponding to 10 nM concentration of caspase-3 is not visible at the shown image intensity however, at an intensity of 19 fluorescent units, the feature is two fold brighter than 25 the background. Thus, for the case of caspase-3, 0.2 pmol of enzyme was sufficient to be detectable with this method.

[132] *Profile of crude cell lysates.* To determine whether PNA-encoded activity based probes were capable of sensitively and specifically measuring differences in protein function in biological samples, an *in vitro* system of cytotoxic lymphocyte mediated 30 cell death was studied. Cytotoxic lymphocytes kill virus-infected or tumor cells through the induction of apoptosis by two contact dependent mechanisms: directed release of granules from the cytotoxic lymphocyte onto the surface of the target cell, and by interaction of the Fas ligand with the Fas receptor (see, Froelich, *et al.*, *J. Biol. Chem.*, 271:29073-29079

(1996)). The predominant mechanism is that of granule release, where the granule protein perforin facilitates the entry of granzyme B, a granule serine proteases, into the cytosol of the targeted cell. Granzyme B then initiates apoptosis primarily through the cleavage-activation of the latent pro-apoptotic cytosolic cysteine protease caspase-3 (see, Nicholson, *et al.*, 5 *Trends Biochem. Sci.*, 22:299-306 (1997)). Molecular regulation of this process occurs at the post-translation level. Indeed, this process is largely independent of protein synthesis and therefore would yield inadequate results by traditional mRNA expression profiling. This process can be conveniently modeled *in vitro* through the incubation of the cytosolic fraction of cellular lysates with granzyme B. Crude cell lysates from Jurkat cells were activated with 10 granzyme B to initiate the apoptosis pathway. As expected based on the selectivity of the probes, incubation of the library with purified granzyme B alone did not give any signal (Figure 16, panel B) whereas incubation with purified caspase-3 showed an intense signal only for the probe corresponding to the caspase-3 inhibitor (Figure 16, panel C). With this negative and positive control at hand, apoptotic crude cell lysates from Jurkat cells was 15 profiled and compare it to the a non-activated sample (Figure 16, panel E and D respectively). The experiments were carried out using 20 L of lysates at a 1 mg/mL concentration (ca. 10<sup>5</sup> cells per profile) with 20 pmol of each probes for 2 hrs. As in the previous experiments, the unbound probes were removed by filtration through a 30 kDa molecular weight cutoff filter and the retained sample was then hybridized to a GenFlex™ 20 oligonucleotide microarray ([www.affymetrix.com](http://www.affymetrix.com)) and directly imaged by fluorescein fluorescence. While the signal corresponding to the cathepsin inhibitors was virtually identical in the two samples, there is a dramatic difference in intensity for the probe corresponding to the caspase-3 inhibitor.

[133] *Target identification and validation.* Most profiling experiments 25 attempt to assign the biochemical origin of a perturbed cellular state by comparing its profile to that of an unperturbed sample. A general issue in such profiling experiments is whether the observed differences in profiles are causal or circumstantial. A unique feature of the approach described here is that the activity of a particular enzyme is measured by the amount of enzyme trapped by a mechanism-based inhibitor on a microarray. Such an inhibitor may 30 be used to isolate the enzyme by affinity chromatography and as an *in vitro* or *in vivo* inhibitor to assess whether there is a correlation between the profile and phenotype. To demonstrate these two points, compound 6, which stood out as a clear difference between apoptotic and nonapoptotic profiles, was used. Thus compound 6b, where the PNA has been substituted for biotin, was incubated with the same apoptotic crude cell lysates that were used

in the profile. The labeled adduct was then immobilized on an avidin resin and washed to remove non-specific adherents. The immobilized protein was then released by incubation with excess biotin and digested with trypsin. The tryptic peptides were analyzed by electrospray ionization mass spectrometry. Tandem mass spectra corresponding to doubly 5 and triply charged SGTDVDAANLRETFR and NKNDLTREEIVELMR peptides were identified with the database searching program SEQUEST (Figure 17). This search confirmed that these peptides could only be derived from caspase-3, validating the affinity capture method (see, Tsaprailis, *et al.*, *J. Am. Chem. Soc.*, 121:5142-5154 (1999)).

[134] Having established a practical protocol to rapidly characterize a protein 10 corresponding to a particular probe, attention was turned to the use of that probe as an inhibitor of the lymphocyte mediated cell death. The lysates were treated with caspase-3 inhibitor 6c (Figure 14) prior to granzyme B activation and apoptosis was measured by monitoring the cleavage of downstream substrates of caspase-3. As shown in Figure 18A, caspase-3 is proteolytically converted by granzyme B to its P20/P12 active enzyme (see, 15 Nicholson, *et al.*, *Nature*, 376:37-43 (1995); and Quan, *et al.*, *Proc. Natl. Acad. Sci. USA*, 93:1972-1976 (1996)) (upper panel), however the proteolytic degradation of DFF-45 (see, Liu, *et al.*, *Cell*, 89:175-184 (1997)) is clearly inhibited by compound 6c (lower panel). It is interesting to note that the covalent caspase-3-inhibitor adduct is detectable on this gel with a band corresponding to 22 KDa. The autoproteolysis of caspase-3 P20 subunit to the mature 20 P17 fragment is also inhibited by 6c. Attention was then turned to a whole cell assay of apoptosis. Jurkat cells were incubated for 12 h with the inhibitor (1  $\mu$ M) prior to the induction of apoptosis with a Fas-activating antibody. The extent of apoptosis was measured using two different stains, Annexin V to measure phosphatidylserine relocation to the extracellular leaflet (early apoptosis) and propidium iodide, a membrane impermeable DNA 25 stain to measure the integrity of the phospholipid bilayer (indicator of late stage apoptosis).

The proportion of apoptotic cells were then measured by fluorescence-activated cell sorting (FACS). Treatment of Jurkat cells with a Fas ligand induced more than 50% apoptosis (Figure 18B, panel 2) relatively to the non treated sample (Figure 18B, panel 1). While the highly charged nature of compound 6c appeared to prevent membrane permeability, a close 30 "prodrug" analogue wherein the aspartic and glutamic acid are methylated (Cbz-Asp(OMe)-Glu(OMe)-Val-Asp(OMe)-FMK) did inhibit this Fas-mediated apoptosis (Figure 18B, panel 3).

[135] In this study, it has been shown that from an observed difference in profile between the apoptotic sample and the non-apoptotic sample the corresponding inhibitor could be used to isolate and identify caspase-3 by affinity capture out of crude cell lysates and characterize it by mass spectrometry. Finally, inhibition of the apoptosis 5 phenotype supports the critical role that caspase-3 plays in apoptosis as revealed in the profile. Although the function of caspase-3 in apoptosis has been extensively studied, it validates the approach presented herein and establishes a working protocol for subsequent discovery work based on this method.

[136] In conclusion, it has been demonstrated that PNA-encoded small 10 molecule microarrays can be a powerful tool to monitor enzymatic activity in a highly miniaturized and parallel format. The methodology to characterize enzymes identified from such small molecule-based microarray has been developed and validated. More importantly, it has been demonstrated that small molecule-based profiling facilitates subsequent chemical biology investigations by providing a small molecule inhibitor to an identified enzyme of 15 interest. While the aim of this study was to validate this small molecule-based profiling approach in a biologically relevant context, it establishes that large combinatorial libraries based on this approach are useful in the discovery of novel enzymes and pathways. It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to 20 persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

**WHAT IS CLAIMED IS:**

1           1. A method for preparing a library of diverse compounds, each of said  
2 compounds being produced by the step-by-step assembly of building blocks, said method  
3 comprising:

4               (a) apportioning solid supports among a plurality of reaction vessels; and  
5               (b) in each reaction vessel of said plurality of reaction vessels, exposing  
6 said solid supports to a first building block of a compound and to a first monomer of a  
7 peptido nucleic acid (PNA) identifier tag that identifies said first building block under  
8 conditions suitable for immobilization of said first building block and said first monomer,  
9 wherein the first building block present in one reaction vessel is different from the first  
10 building block present in at least one of the other reaction vessels, wherein said first building  
11 block of said first compound is capable of being covalently coupled to a second building  
12 block and wherein said first monomer of said PNA identifier tag is capable of being  
13 covalently coupled to a second monomer.

1           2. The method in accordance with claim 1, further comprising:  
2               (c) pooling said solid supports.

1           3. The method in accordance with claim 2, further comprising:  
2               (d) reapportioning the pooled solid supports among a plurality of reaction  
3 vessels; and

4               (e) in each reaction vessel of said plurality of reaction vessels, exposing  
5 said solid supports to at least a second building block of said compound and to at least a  
6 second monomer of said PNA identifier tag under conditions suitable for attachment of said  
7 second building block to said first building block and said second monomer to said first  
8 monomer, wherein said second building block present in one reaction vessel is different from  
9 said second building block present in at least one of the other reaction vessels.

1           4. The method in accordance with claim 1, wherein said solid supports  
2 that are apportioned in (a) each further comprise at least a third building block of the  
3 compound and at least a third monomer of the PNA identifier tag, wherein the third monomer  
4 of the PNA identifier tag identifies the third building block, and wherein said first building  
5 block attaches to said third building block and said first monomer attaches to said third  
6 monomer.

1               5.     The method in accordance with claim 1, wherein said first building  
2     block of said compound is an amino acid.

1               6.     The method in accordance with claim 3, wherein said amino acid is a  
2     member selected from the group consisting of L-amino acids, D-amino acids,  $\alpha$ -amino acids,  
3      $\beta$ -amino acids and  $\omega$ -amino acids.

1               7.     The method in accordance with claim 1, wherein said PNA identifier  
2     tag is from about 3 to about 50 nucleotides in length.

1               8.     The method in accordance with claim 5, wherein said PNA identifier  
2     tag is from about 6 to about 20 nucleotides in length.

1               9.     The method in accordance with claim 5, wherein said PNA identifier  
2     tag is about 12 nucleotides in length.

1               10.    The method in accordance with claim 5, wherein said monomer of said  
2     PNA identifier tag comprises 2 or more nucleotides.

1               11.    The method in accordance with claim 5, wherein said monomer of said  
2     PNA identifier tag comprises 5 or fewer nucleotides.

1               12.    The method in accordance with claim 1, wherein said PNA identifier  
2     tag further comprises a label.

1               13.    The method in accordance with claim 12, wherein said label is a  
2     fluorophore.

---

1               14.    The method in accordance with claim 12, wherein said label is a  
2     radioactive label.

1               15.    The method in accordance with claim 1, wherein said first building  
2     block is immobilized on said solid support.

1               16.    The method in accordance with claim 1, wherein said first monomer is  
2     immobilized on said solid support.

1                   17. The method in accordance with claim 1, wherein said first monomer is  
2 immobilized on said first building block and not said solid support.

1                   18. The method in accordance with claim 1, wherein said first building  
2 block is immobilized on said solid support and said first monomer is immobilized on said  
3 first building block.

1                   19. The method in accordance with claim 13, wherein said first monomer  
2 is immobilized on said first building block through a linker.

1                   20. The method in accordance with claim 1, wherein said solid support is a  
2 bead or particle.

1                   21. The method in accordance with claim 1, wherein said solid support is a  
2 nonporous bead.

1                   22. The method in accordance with claim 1, wherein said solid support is a  
2 bead having a diameter ranging from about 1 nm to about 1 mm.

1                   23. The method in accordance with claim 1, wherein prior to exposing said  
2 first building block to said solid support, said first building block is activated to facilitate  
3 immobilization of said first building block onto said solid support.

1                   24. The method in accordance with claim 1, wherein prior to exposing said  
2 first monomer to said solid support, said first monomer is activated to facilitate  
3 immobilization of said first monomer onto said solid support.

1                   25. The method in accordance with claim 1, wherein prior to exposing said  
2 first monomer to said solid support, said first monomer is activated to facilitate  
3 immobilization of said first monomer onto said first building block.

1                   26. The method in accordance with claim 1, wherein said solid support is  
2 exposed to said first monomer after said solid support is exposed to said first building block.

1                   27. The method in accordance with claim 1, further comprising:  
2                   (c) cleaving said compound from said solid support.

1                   28. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least 10 different compounds.

1                   29. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least 100 different compounds.

1                   30. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least  $10^3$  different compounds.

1                   31. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least  $10^4$  different compounds.

1                   32. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least  $10^5$  different compounds.

1                   33. The method in accordance with claim 3, wherein steps (a) through (e)  
2 are carried out so as to construct a library of at least  $10^6$  different compounds.

1                   34. A method for identifying a compound that binds a target, said method  
2 comprising:

3                   (a) contacting said target with a library of compounds, wherein each of  
4                   said compounds comprises a peptido nucleic acid (PNA) identifier tag;

5                   (b) separating the compounds that bind said target from those compounds  
6                   that do not bind said target to obtain target-compound complexes;

7                   (c) hybridizing the target-compound complexes to an array of  
8                   oligonucleotides; and

9                   (d) detecting the target-compound complexes that hybridize to said array

10                   of oligonucleotides, thereby identifying said compounds that bind said  
11                   target.

1                   35. The method in accordance with claim 34, wherein said target is a  
2 protein.

1                   36. The method in accordance with claim 34, wherein said target is in a  
2 cell extract, tissue or other biological sample.

1                   37.    The method in accordance with claim 34, wherein said target  
2   comprises a label.

1                   38.    The method in accordance with claim 37, wherein said label is a  
2   fluorophore.

1                   39.    The method in accordance with claim 34, wherein each of said  
2   compounds further comprises a label.

1                   40.    The method in accordance with claim 34, wherein said label is attached  
2   to said PNA identifier tag.

1                   41.    The method in accordance with claim 40, wherein said label is a  
2   fluorophore.

1                   42.    The method in accordance with claim 34, wherein (b) is carried out  
2   using size-exclusion chromatography.

1                   43.    The method in accordance with claim 34, wherein said target is a  
2   library of targets.

1                   44.    The method in accordance with claim 34, wherein each of the  
2   oligonucleotides in said array is about 10 to about 50 nucleotides in length.

1                   45.    The method in accordance with claim 34, wherein each of the  
2   oligonucleotides in said array is about 20 to about 30 nucleotides in length.

1                   46.    The method in accordance with claim 34, wherein the PNA identifier  
2   tag hybridizes to a terminal portion of the oligonucleotide.

1                   47.    A method for identifying a compound that binds a target, said method  
2   comprising:  
3                   (a)   providing a library of compounds, wherein each of said compounds  
4                   comprises a peptido nucleic acid (PNA) identifier tag;  
5                   (b)   hybridizing said library of compounds to an array of oligonucleotides;  
6                   (c)   contacting said array of bound compounds with a target; and  
7                   (d)   determining which compounds bind said target.



1                   48. The method in accordance with claim 47, wherein said target is a  
2 protein.

1                   49. The method in accordance with claim 47, wherein said target  
2 comprises a label.

1                   50. The method in accordance with claim 49, wherein said label is a  
2 fluorophore.

1                   51. The method in accordance with claim 47, wherein said target is a  
2 library of targets.

1                   52. The method in accordance with claim 51, wherein each of said targets  
2 comprises a different label.

1                   53. The method in accordance with claim 47, wherein (c) is performed  
2 before (b).

1                   54. The method in accordance with claim 47, wherein said compounds  
2 comprise one or more labels, and said determining which targets bind said compound  
3 comprises determining whether the label is removed from the compound by contact with the  
4 target.

1                   55. The method in accordance with claim 54, wherein said compounds  
2 each comprise two labels, and said determining which targets bind said compound comprises  
3 determining whether either or both label is removed from the compound by contact with the  
4 target.

---

1                   56. The method in accordance with claim 55, wherein said compounds  
2 each comprise two fluorescent labels, and said determining which targets bind said  
3 compound comprises detecting a change in fluorescence resonance energy transfer (FRET)  
4 that results from contact with the target.

Figure 1

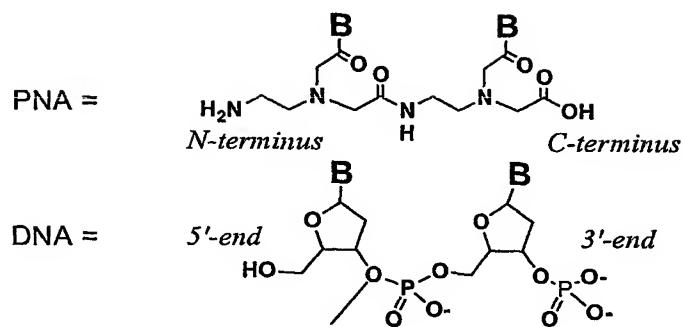


Figure 2

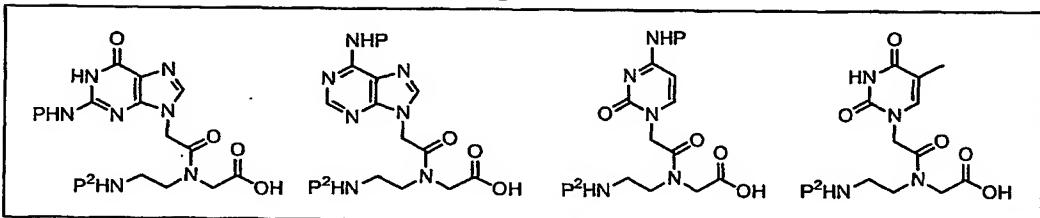


Figure 3

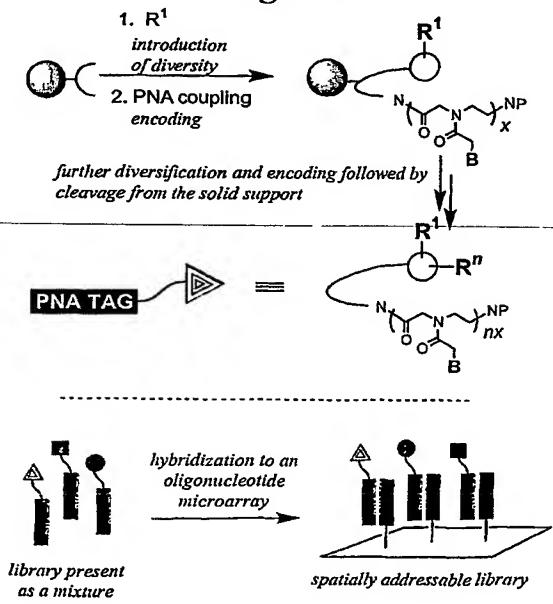


Figure 4

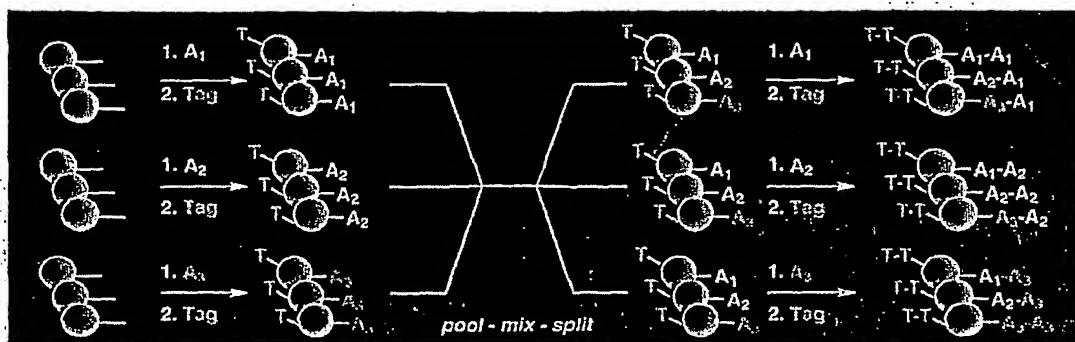


Figure 5

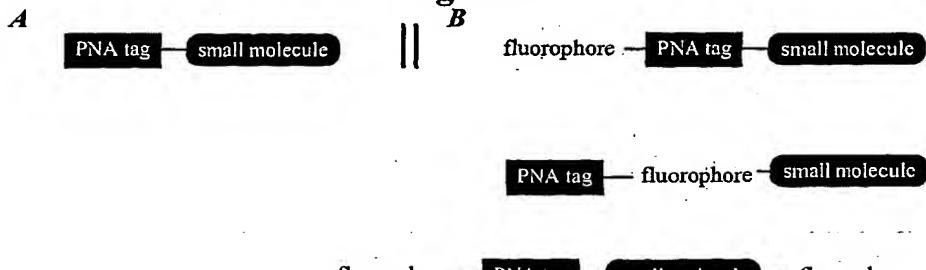


Figure 6

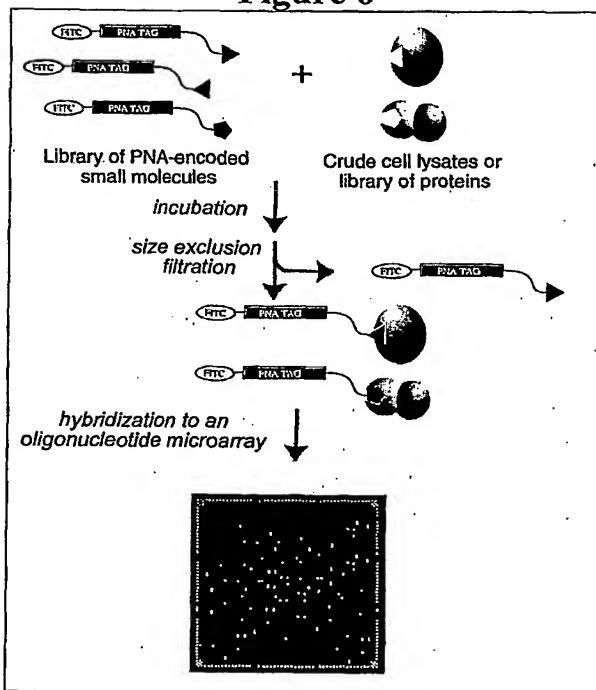


Figure 7

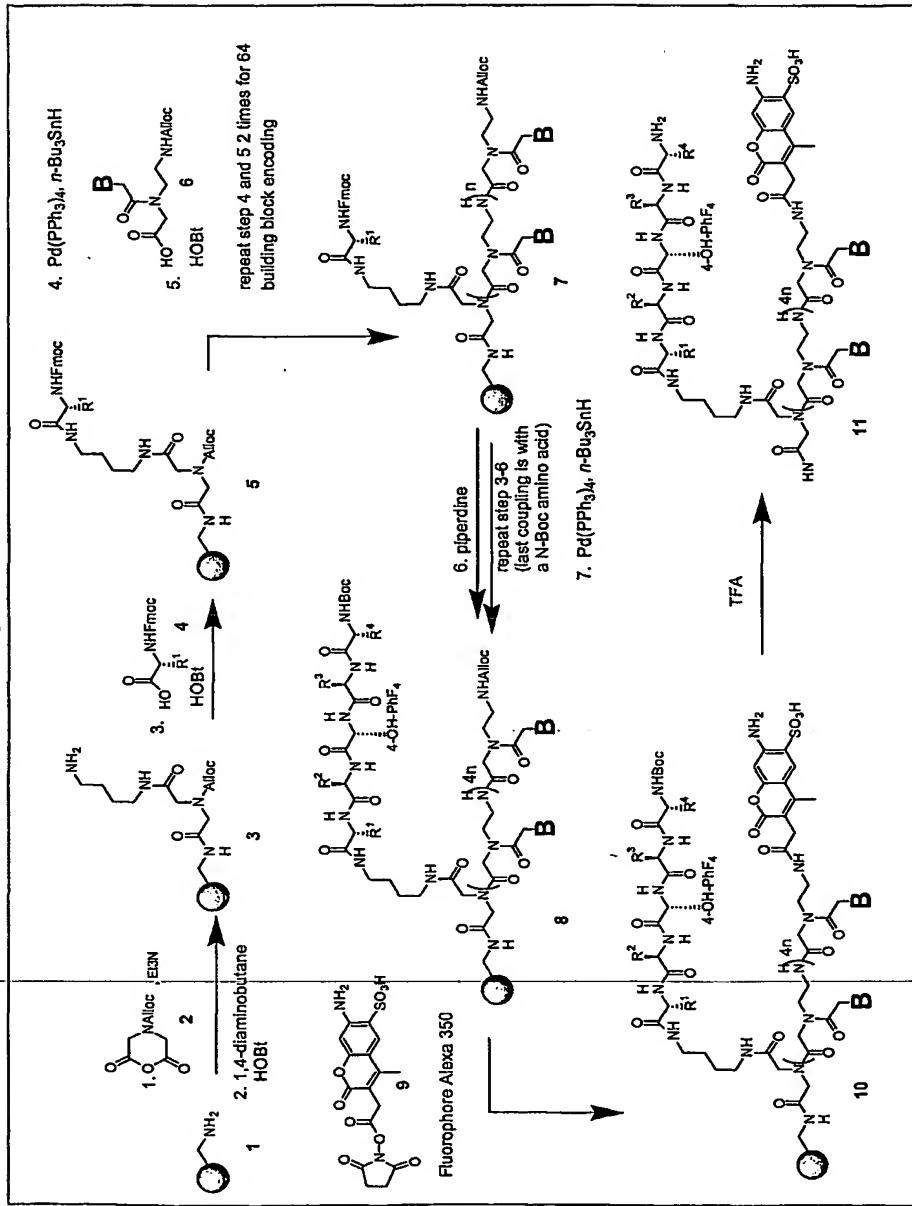


Figure 8

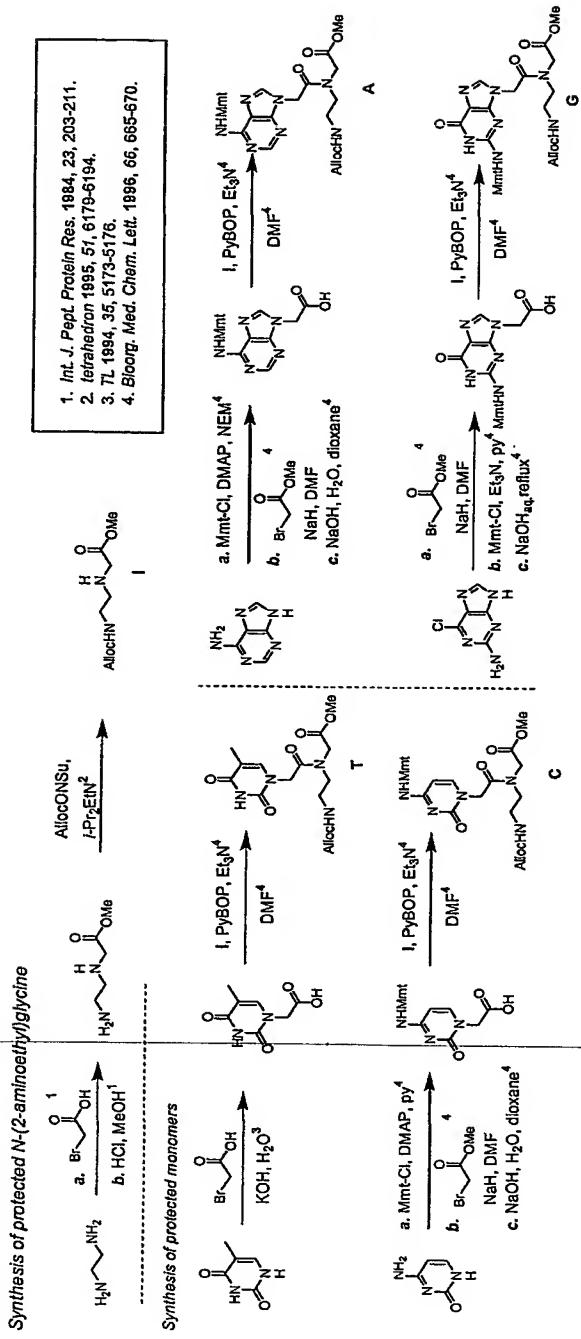


Figure 9

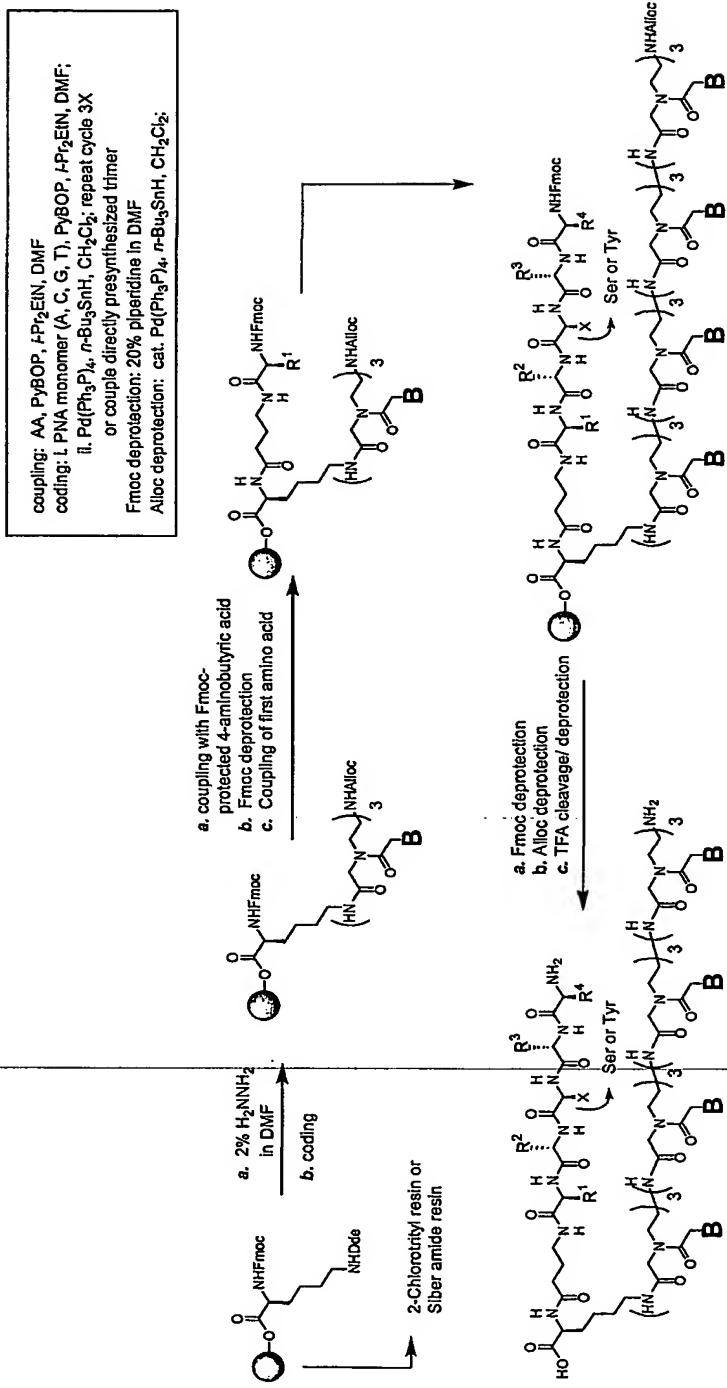


Figure 10

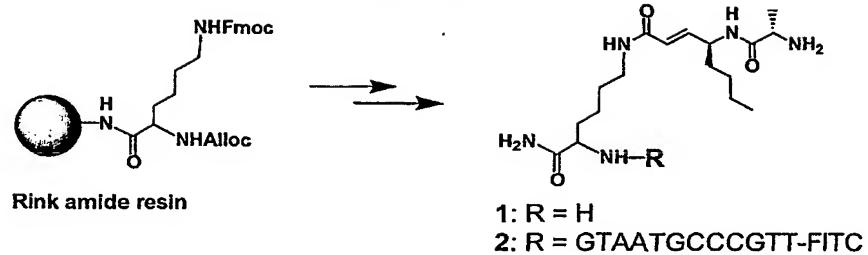


Figure 11

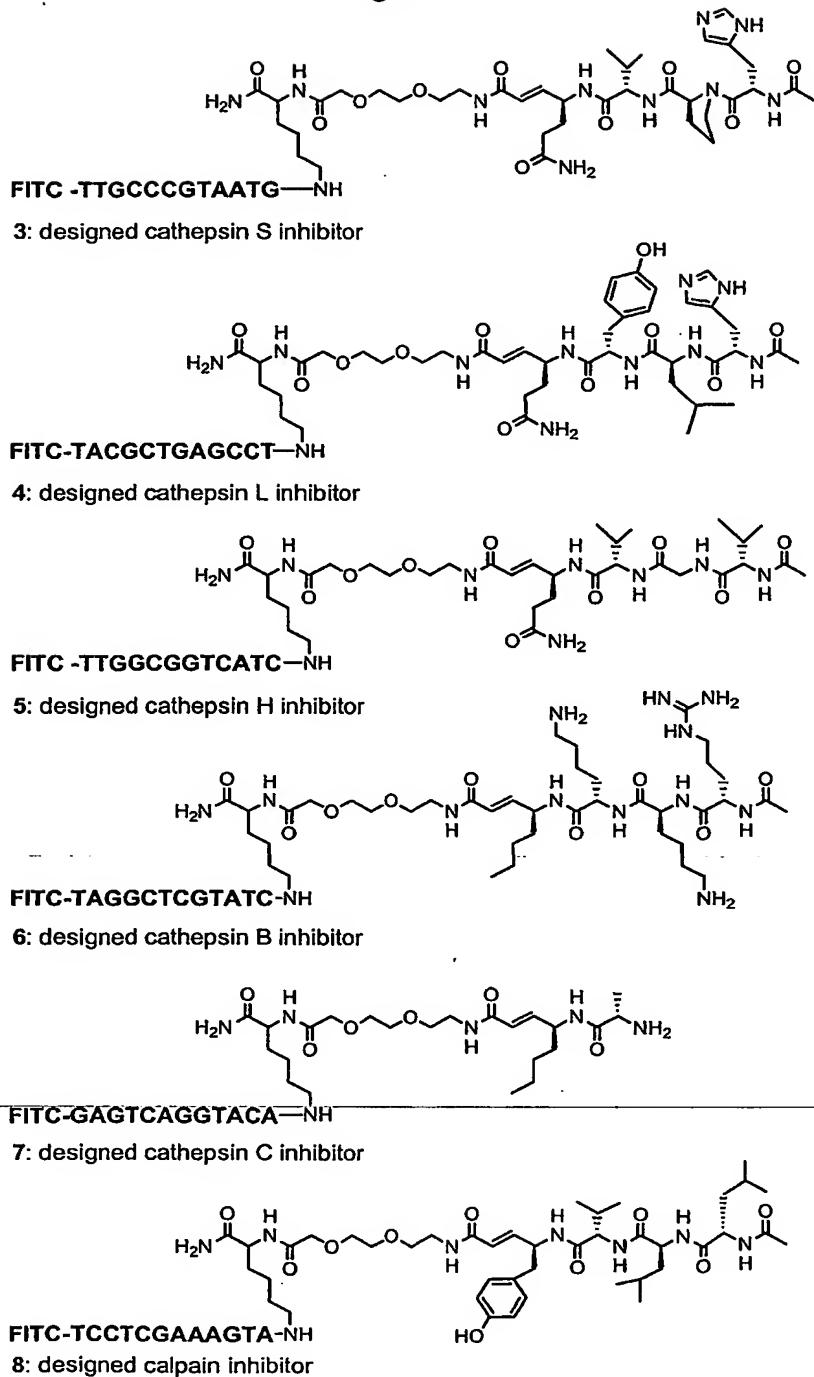


Figure 12

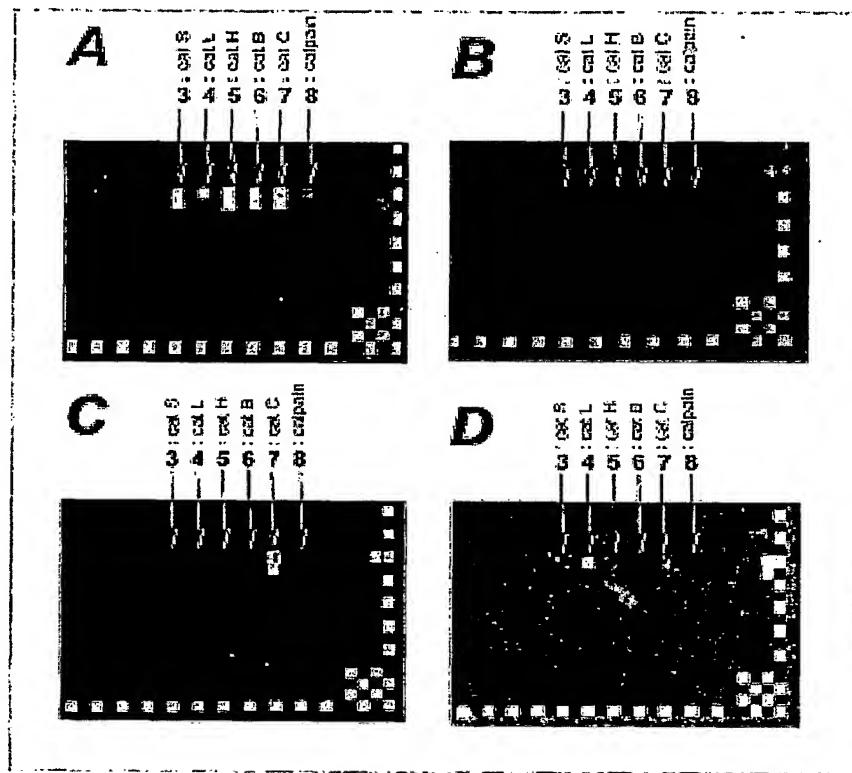
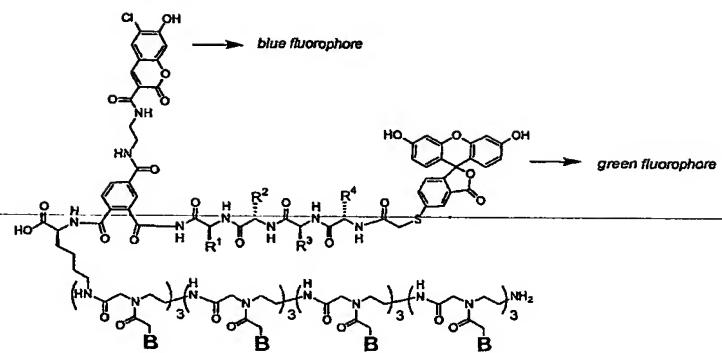
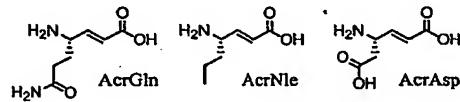
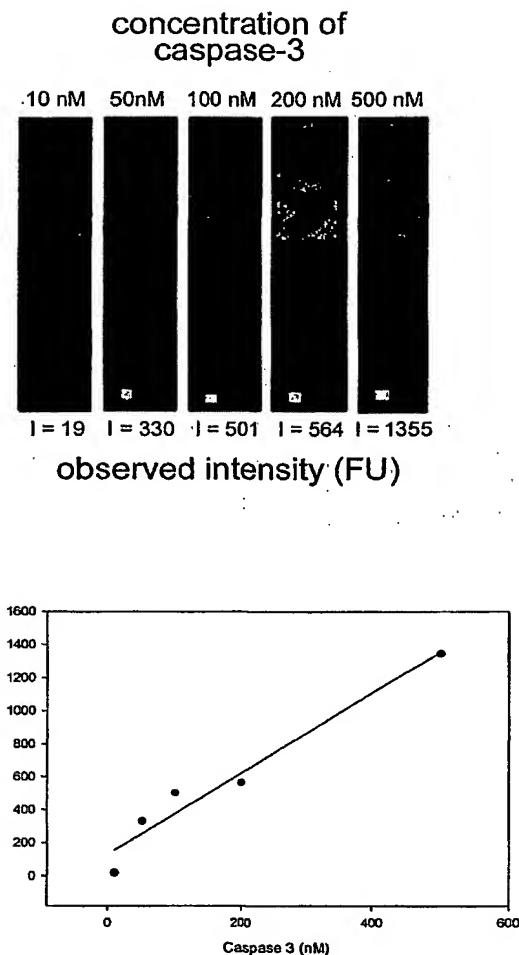


Figure 13



**Ac-His-Pro-Val-AcrGln-PEG-Lys(PNA-FITC)NH<sub>2</sub>****1:** designed cathepsin S inhibitor**Ac-His-Leu-Tyr-AcrGln-PEG-Lys(PNA-FITC)NH<sub>2</sub>****2:** designed cathepsin L inhibitor**Ac-Wal-Gly-Val-AcrGln-PEG-Lys(PNA-FITC)NH<sub>2</sub>****3:** designed cathepsin H inhibitor**Ac-Arg-Lys-Lys-AcrNle-PEG-Lys(PNA-FITC)NH<sub>2</sub>****4:** designed cathepsin B inhibitor**H<sub>2</sub>NAla-AcrNle-PEG-Lys(PNA-FITC)NH<sub>2</sub>****5:** designed cathepsin C inhibitor**Ac-Asp-Val-Glu-AcrAsp-PEG-Lys(R)NH<sub>2</sub>****6a:** R = PNA-FITC, designed caspase-3 inhibitor**6b:** R = LC-biotin, caspase-3 affinity label**6c:** R = H, caspase-3 inhibitor**Ac-Asp-Val-Glu-AccAsp-PEG-Lys(PNA-FITC)NH<sub>2</sub>****7:** designed cathepsin K inhibitor**Fig. 14**



**Fig. 15**

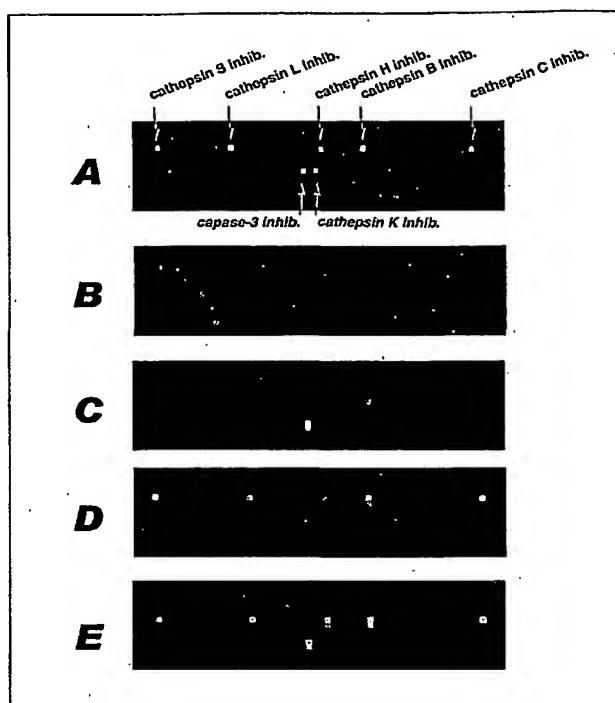


Fig. 16

13/15

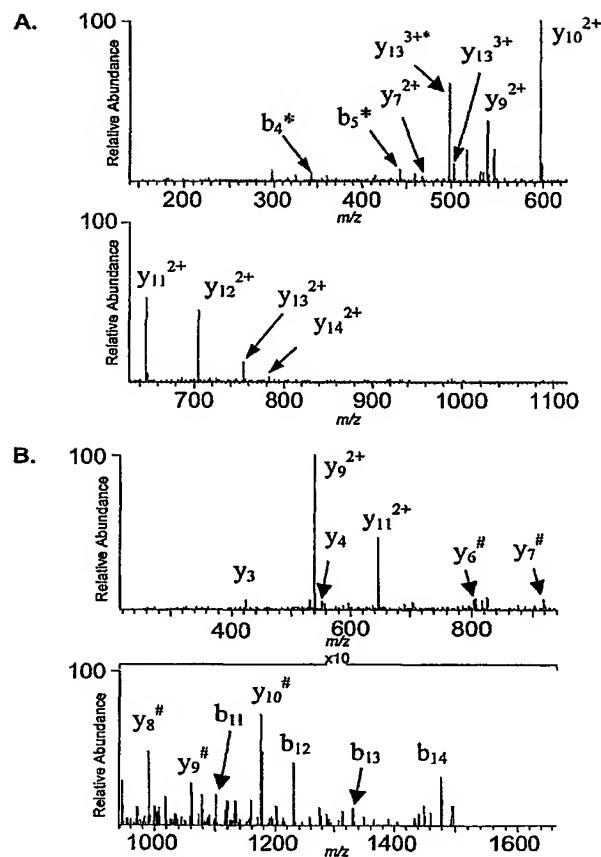
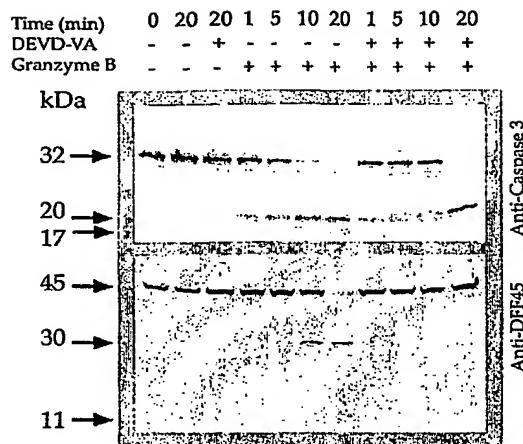
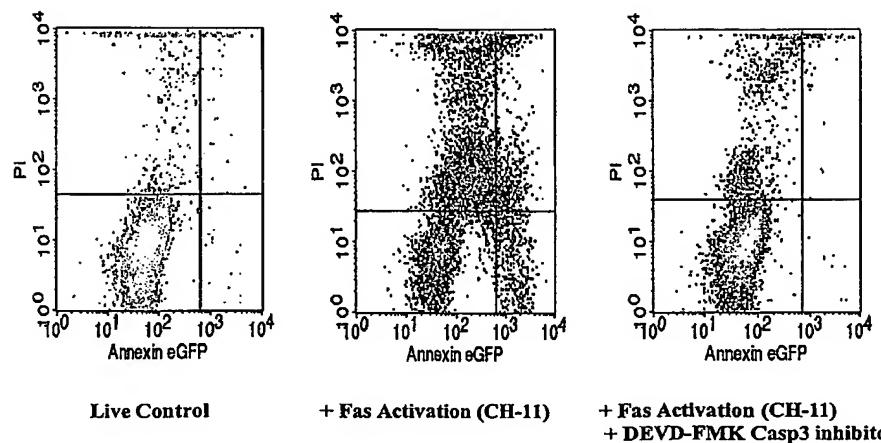


Fig. 17

**A.****B.****Fig. 18**

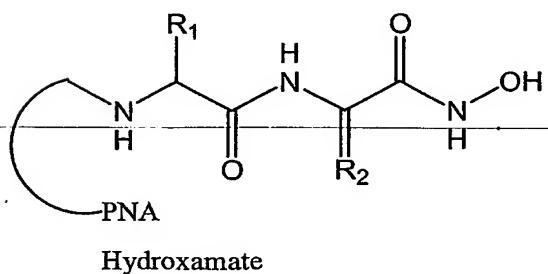
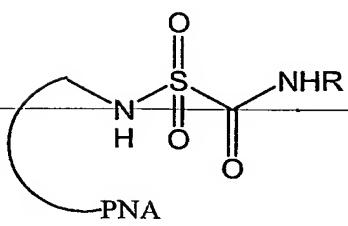
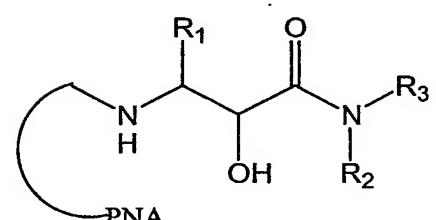
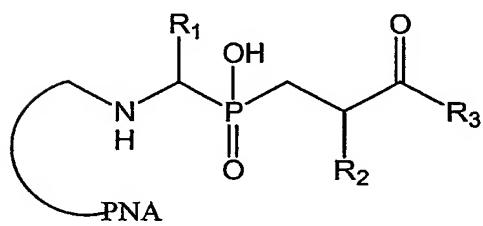
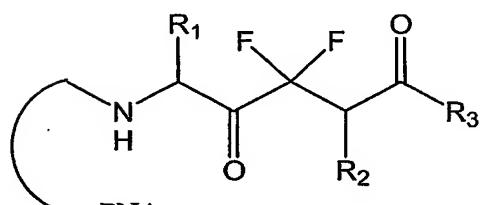
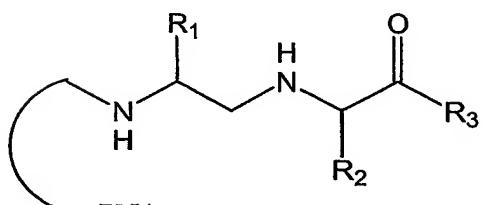
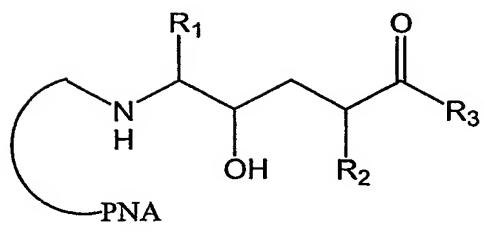
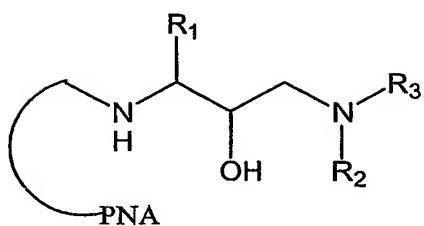


Fig. 19

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number  
WO 02/099078 A3

(51) International Patent Classification<sup>7</sup>: C12Q 1/68, (81) Designated States (national): AE, AG, AL, AM, AT, AU, C07H 21/04, 21/02 AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US02/18065

(22) International Filing Date: 5 June 2002 (05.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/296,525 5 June 2001 (05.06.2001) US  
60/363,901 11 March 2002 (11.03.2002) US

(71) Applicant (for all designated States except US): IRM LLC  
[US/US]; P.O. Box HM 2899, HM LX Hamilton (BM).

(72) Inventors; and

(75) Inventors/Applicants (for US only): WINSSINGER, Nicolas [BE/US]; 643 Gravilla Street, La Jolla, CA 92037 (US). HARRIS, Jennifer, L. [US/US]; 4572 Tivoli Street, San Diego, CA 92107 (US). BACKES, Bradley, J. [US/US]; 1901A Oxford Avenue, Cardiff by the Sea, CA 92007 (US). SCHULTZ, Peter, G. [US/US]; 1650 La Jolla Rancho, La Jolla, CA 92037 (US).

(74) Agents: GARRETT-WACKOWSKI, Eugenia et al.; TOWNSEND AND TOWNSEND AND CREW LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— with international search report  
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:  
6 March 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/099078 A3

(54) Title: FUNCTIONAL PROTEOMIC PROFILING

(57) Abstract: Oligonucleotides such as PNA are used to code for the identity of individual members within a library such that a library present as a mixture can be converted into a spatially addressable format through hybridization to an oligonucleotide array. Such a strategy can greatly facilitate arraying small molecules, antibodies, proteins and oligosaccharides on a chip and allows for binding or other assays to be performed in solution prior to hybridization. This invention is particularly useful for chemical libraries as it allows for combinatorial synthesis employing split and mix technology with positional encoding. The decoding is achieved by hybridization to an oligonucleotide. With high density oligonucleotide arrays, every library member can be analyzed in a highly miniaturized format. As such, this technology readily lends itself to highly miniaturized screening of single or multiple targets simultaneously and profiling.

## INTERNATIONAL SEARCH REPORT

International application No:  
PCT/US02/18065

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/04, 21/02;  
US CL : 435/6; 536 23.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

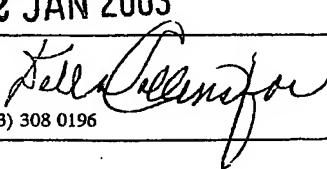
Minimum documentation searched (classification system followed by classification symbols)  
U.S. : 435/6; 536 23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
WEST 2.0, CAPLUS, MEDLINE, BIOSIS, EMBASE, BIOTECHDS, SCISEARCH

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	THOMPSON et al. Synthesis and Applications of Small Molecule Libraries. Chem. Rev. December 1996, Vol. 96, No. 1, pages 555-600, see entire reference.	1-56
Y	WO 00/07017 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 10 February 2000 (02.10.2000) entire patent especially pages 5-41.	1-56
Y	US 6,027,890 A (NESS et al.) 22 February, 2000 (02.22.2000), entire patent, especially columns 15-42.	1-56
Y	SCHENA et al. Quantitative Monitoring of gene expression patterns with a complementary DNA microarray. Science. October 1995, Vol. 270, pages 467-470, especially page 467-468.	34-56

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 03 November 2002 (03.11.2002)		Date of mailing of the international search report 02 JAN 2003	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230		Authorized officer Gary Benzion Telephone No. (703) 308 0196 	

Form PCT/ISA/210 (second sheet) (July 1998)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**